The Dissertation Committee for Amber Chen Certifies that this is the approved version of the following dissertation:

Subnetwork Analysis: Methodology and Application

Committee:

Randy Machemehl, Supervisor

Zhanmin Zhang

Stephen Boyles

Fernanda Leite

Ming Zhang

Subnetwork Analysis: Methodology and Application

by

Amber Chen

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2017

ACKNOWLEDGEMENTS

I would like to first thank Dr. Randy Machemehl for his guidance and support during my graduate study and my doctorate in civil engineering. Dr. Machemehl and the other faculty and staff at the University of Texas at Austin have taught and led me to become a better researcher. I am grateful that Dr. Randy Machemehl allowed me to be a teaching assistant for his undergraduate course. I am grateful for Carolina Baumanis's help in revising my papers and dissertation. I am also thankful that Lisa Smith acclimated me to the new American culture and helped me improve English in the beginning of my graduate study.

I also want to thank my committee: Dr. Stephen Boyles, Dr. Zhanmin Zhang, Dr. Fernanda Leite, and Dr. Ming Zhang for their advice on a wide range of topics from dissertation writing to career development. I owe a special thank you to Dr. Boyles and Dr. Amit Bhasin for generously allowing me to be a teaching assistant for their undergraduate course. I would like to acknowledge Dr. Tsai-Yun Liao who provided her support, and research and teaching experience in UT. I would also like to thank Dr. Kara Kockeman who helped me apply for the IGERT fellowship. The following list of other friends and colleagues have guided or assisted me through graduate school: Prateek Bansal, Amit Singh, Jiangling Wu, Dr. Moggan Motamed, and Dr. Mubassira Khan. Without their insight and advice I would not be where I am today

Additionally, I would also like to thank the Network Modeling Center team at the Center for Transportation Research, notably Dr. Mason Gemar, Dr. Natalia Ruiz Juri, Dr. Dennis Bell, and Dr. Itamar Gal, who provided invaluable input, guidance, and support for this research effort.

Finally, I would like to thank my parents, Chia-Hua Chen and Li-Ching Lien, and my brother, Pon-Hou Chen for their love and support. I would like to extend my gratitude to my friends and classmates. They have all been there for me as I completed this dissertation, especially my fiancé, Ernie Hwaun, who with his love and patience saw me through to the end.

Subnetwork Analysis: Methodology and Application

Amber Chen, Ph.D.

The University of Texas at Austin, 2017

Supervisor: Randy Machemehl

The focus of this dissertation is to create robust tools that enable efficient and comprehensive subnetwork analysis for Dynamic Traffic Assignment (DTA) and a microscopic simulation setting. A DTA subnetwork can potentially replace a large urban transportation network that experiences a change in only a small fraction of the whole network. However, DTA mainly uses Cell Transmission Model (CTM), which lacks many details provided through microscopic traffic simulation. Also, there is very little research done on the balance between the computational time and the subnetwork size. Computational time increases when using a larger subnetwork, but the simulated result is more similar to that of the entire network. Conversely, the computational time decreases when using a smaller subnetwork, but the simulated result might not replicated the entire network.

Currently, extracting a subnetwork is a manual and time-consuming process, requiring an entire coded urban network in ArcGIS. Therefore, to overcome these shortcomings this study automated the process of extracting a subnetwork. Moreover, to further the transition between long-term and short-term traffic analysis, the study integrated a DTA simulator and a microscopic traffic simulator so that together they can assign traffic and provide detailed traffic result. This study also defined an appropriate sub-arterial size for the microscopic simulator, which is not the same as the size of the DTA subnetwork. Furthermore, this study analyzed several factors which significantly influence

computational time, and developed optimization models to find the balance between the computational time and error resulting from sub-area size.

Ultimately, this study developed two programs that can automatically extract a subnetwork from a regional DTA network, and automatically develop an identical subnetwork in a microscopic simulator from this DTA network of an appropriate size. The methodologies this study built promote the efficient analysis of traffic conditions and facilitate the implementation of advanced models that were previously limiting in terms of the amount of time required to compute results; also, the automatic tools this study developed will contribute to the depth and the breadth of dynamic transportation systems analysis.

LIST OF CONTENTS

ACKNOWLEDGEMENTS ..... **I**

ABSTRACT ..... **VI**

LIST OF TABLES ..... **XI**

LIST OF FIGURES ..... **XIII**

CHAPTER 1: INTRODUCTION **1**

1.1 Background 2

1.2 Motivation 5

1.3 Research Objectives and Contribution 8

1.4 Data 10

1.5 Overview 11

CHAPTER 2: LITERATURE REVIEW **13**

2.1 Traffic Assignment 13

2.1.1 Static Traffic Assignment (STA) 15

2.1.2 Dynamic Traffic Assignment (DTA) 17

2.1.3 STA versus DTA 19

2.2 Subnetwork Analysis 21

2.2.1 Sub network and sub OD Matrix Development 23

2.2.2 Subnetwork Evaluation 25

2.3 DTA program: VISTA and MatSIM 29

2.3.1 VISTA 29

2.3.2 MATSIM (Multi-Agent Transportation Simulation) 30

2.4 Microsimulators: CORSIM and VISSIM 32

2.5 CPU Performance 36

2.6 Summary 37

CHAPTER 3. DATA 39

CHAPTER 4. OPTIMAL SUB-AREA SIZE AND SIMULATION TIME 44

4.1 Simulation Time Estimation 45

4.2. Objective Functions and Constraints 54

4.3 Case Study 56

4.4 Summary 59

CHAPTER 5. LINEAR REGRESSIONS FOR CORSIM SUB-ARTERIAL SIZE.....61

5.1 Methodology 61

5.2 Statistical Analysis 64

5.3 Boxplot Analysis 68

5.4 Models 107

5.4.1 Impact RMSE Model 110

5.4.2 Confidence Interval Model 113

5.4.3 Result and Analysis 114

CHAPTER 6. DTA SUBAREA DEVELOPER 123

6.1 First subsystem.....125

6.2 Second subsystem 125

6.3 Case Study 126

6.4 Summary.....128

CHAPTER 7 CORSIM TRANSFORMER 130

7.1 Developing an identical network in CORSIM.....130

7.2 Result 144

7.3 Summary.....145

CHAPTER 8 FINAL 147

8.1 Summary.....147

8.2 Contribution 148

8.3 Future Research 149

APPENDIX.....151

REFERENCES.....156

LIST OF TABLES

Table 1 Example of the reduction of simulation time in Dallas-Fort Worth network (Gemar, 2013) 22
Table 2 Example of the reduction of file space required for Dallas-Fort Worth network (Gemar, 2013) 22
Table 3 shows the number of the links for the sub-area size and work zone sizes in Guadalupe, 7th, and West MLK. 39
Table 4 The individual and average work zone traffic flow across the sub-area size, work zone size, and networks 42
Table 5 The correlations between the indexes and simulation time 50
Table 6 Statistical result of predicting simulation time using demand 51
Table 7 Statistical result of predicting simulation time using zone and size 52
Table 8 Example result for Z1 objective function 57
Table 9 Example result for Z2 objective function 58
Table 10 Statistical Tests of the RMSE for 10 Base and 10 Impacted Scenarios (Simulation time = 3600 sec) 65
Table 11 Model Specifications Tested Using Link Variables 111
Table 12 Results from the Final Impact Model 113
Table 13 Comparison of the Final Impact Model and Confidence Interval Model for networks with a (7, 1) configuration 115
Table 14 Comparison of the Final Impact Model and Confidence Interval Model for networks with a (9, 2) configuration 117
Table 15 Comparison of the Final Impact Model and Confidence Interval Model for networks with a (11, 3) configuration 119
Table 16 Comparison of the Final Impact Model and Confidence Interval Model for network configurations, including capacity reduction, sub-area size, and work zone size. 122
Table 17 Number of links needed based on size order for 15th, 7th, and Guadalupe Street 128

Table 18 Example of node position data from VISTA 133
Table 19 Example of link detail data from VISTA 133
Table 20 Example of the N-by-N connected node matrix 135
Table 21 Example of vehicle path data 139
Table 22 Example of vehicle path time data 139
Table 23 Example of signal offset data from VISTA 143
Table 24 Example of signal timing data from VISTA 143
Table 25 The ratio of travel time difference between VISTA and CORSIM 145

LIST OF FIGURES

Figure 1 a) Map from Google Maps of the area and b) Example of an elevated crossroad misrepresented as an at-grade intersection in ArcGIS.. 6
Figure 2 a) Base scenario using Guadalupe Street subnetwork and b) 25% capacity reduction on Guadalupe Street c) 50% capacity reduction on Guadalupe Street d) 75% capacity reduction on Guadalupe Street. 8
Figure 3 Primary steps of traffic assignment 13
Figure 4 Example of subnetwork in the entire network 21
Figure 5 Visualization of the Subnetwork Selection with size parameter Process (from Gemar, 2013) 23
Figure 6 Schematic of the departure time and arrival time in the subnetwork 25
Figure 7 Recommended Subnetwork Sizes (from Gemar, 2013) 27
Figure 8 The Downtown Austin network in VISTA 30
Figure 9 MatSIM demand sample code and comments 31
Figure 10 The locations of the work zones and the arterials (a) Guadalupe Subnetwork (b) 7th Street Subnetwork (c) West MLK Subnetwork 41
Figure 11 Example of the grid and the radburn network pattern (The left is the grid, and the right the is radburn pattern) 45
Figure 12 Relationship between demand and simulation time 46
Figure 13 Relationship between number of links link and simulation time 47
Figure 14 Relationship between size order and simulation time 47
Figure 15 Relationship between zone (number of impacted links) and simulation time.. 48
Figure 16 Relationship between zone capacity reduction and simulation time 49
Figure 17 shows an example of the order flow (modified from Gemar, 2013) 63
Figure 18 The chosen links in blue for the boxplots (a) Guadalupe Street, (b) 7th Street, and (c) West MLK. The work zone is shown in red. 69
Figure 19 Boxplots for four selected links in the Guadalupe Street Network with a (7, 1) configuration undergoing capacity reduction.. 70

Figure 20 Boxplots for four selected links in the 7th Street Network with a (7, 1) configuration undergoing capacity reduction.. 72
Figure 21 Boxplots for four selected links in the West MLK Network with a (7, 1) configuration undergoing capacity reduction.. 74
Figure 22 Boxplots for four selected links in the Guadalupe Network with a (9, 2) configuration undergoing capacity reduction.. 75
Figure 23 Boxplots for four selected links in the 7th Street Network with a (9, 2) configuration undergoing capacity reduction.. 76
Figure 24 Capacity reduction boxplots for four selected links in the West MLK Network with a (9, 2) configuration.. 78
Figure 25 Boxplots for four selected links in the Guadalupe Network with a (11, 3) configuration undergoing capacity reduction.. 79
Figure 26 Boxplots for four selected links in the 7th Street Network with a (11, 3) configuration undergoing capacity reduction.. 80
Figure 27 Boxplots for four selected links in the West MLK Network with a (11, 3) configuration undergoing capacity reduction.. 82
Figure 28 Boxplots for four selected links in the Guadalupe Network with a (7, 25%) configuration with varying work-zone sizes. 84
Figure 29 Boxplots for four selected links in the 7th Street Network with a (7, 25%) configuration with varying work-zone sizes..... 85
Figure 30 Boxplots for four selected links in the West MLK Network with a (7, 25%) configuration with varying work-zone sizes..... 86
Figure 31 Boxplots for four selected links in the Guadalupe Street Network with a (9, 50%) configuration with varying work-zone sizes. 88
Figure 32 Boxplots for four selected links in the 7th Street Network with a (9, 50%) configuration with varying work-zone sizes..... 89
Figure 33 Boxplots for four selected links in the West MLK Network with a (9, 50%) configuration with varying work-zone sizes..... 90

Figure 34 Boxplots for four selected links in the Guadalupe Network with a (11, 75%) configuration with varying work-zone sizes. 91
Figure 35 Boxplots for four selected links in the 7th Network with a (11, 75%) configuration with varying work-zone sizes..... 92
Figure 36 Boxplots for four selected links in the West MLK Network with a (11, 75%) configuration with varying work-zone sizes..... 93
Figure 37 Boxplots for four selected links in the Guadalupe Street Network with a (1, 25%) configuration with varying sub-area size.. 96
Figure 38 Boxplots for four selected links in the 7th Street Network with a (1, 25%) configuration with varying sub-area size..... 97
Figure 39 Boxplots for four selected links in the West MLK Network with a (1, 25%) configuration with varying sub-area size..... 98
Figure 40 Boxplots for four selected links in the Guadalupe Street Network with a (2, 50%) configuration with varying sub-area size.. 99
Figure 41 Boxplots for four selected links in the 7th Street Network with a (2, 50%) configuration with varying sub-area size..... 100
Figure 42 Boxplots for four selected links in the West MLK Network with a (2, 50%) configuration with varying sub-area size..... 101
Figure 43 Boxplots for four selected links in the Guadalupe Street Network with a (3, 75%) configuration with varying sub-area size.. 103
Figure 44 Boxplots for four selected links in the 7th Street Network with a (3, 75%) configuration with varying sub-area size. 104
Figure 45 Boxplots for four selected links in the West MLK Network with a (3, 75%) configuration with varying sub-area size..... 105
Figure 46 Example of a sub-area and a target link order..... 110
Figure 47 The modified subnetwork program inputs 124
Figure 48 Automatic Program Procedure 125
Figure 49 Example of a subnetwork developed with the automatic program. To the left, the entire network and to the right, the subnetwork..... 126

Figure 50 Case study locations (Bringardner, 2015) 127
Figure 51 Example of CORSIM tno file 131
Figure 52 Example of CORSIM trf file 132
Figure 53 Example of upstream, downstream, and downstream departure nodes in CORSIM 134
Figure 54 Example of CORSIM's entry flow from the boundary or inside of the network 136
Figure 55 Example of the CORSIM network in animated form 136
Figure 56 Example of centroids and connectors in VISTA 137
Figure 57 Example of RT 50 in trf files 138
Figure 58 Example of multi-connectors to a centroid in VISTA 140
Figure 59 Example of multi-boundary nodes in CORSIM from a VISTA centroid 140
Figure 60 Example of RT 11 and RT 21 in trf files 141
Figure 61 CORSIM's sign and pre-timed signal control codes 142
Figure 62 Example of the RT 35 and RT 36 entry codes in the trf file 143
Figure 63 The transformed network by using CORSIM Transformer 144
Figure 64 Code for extracting sub-area's links and nodes from an entire network (Language: C++). 152
Figure 65 Code for defining the subnetwork in CORSIM (Language: C++). 153
Figure 66 Code for transforming from vehicle paths to link volumes (Language: C++). 154
Figure 67 Code for optimizing sub-area size (Language: GAMS) 155

CHAPTER 1: INTRODUCTION

The traditional approach to assessing impacts of improvements to urban transportation systems assumes that current measured traffic demand is unchanging. That is, increasing the capacity of a congested intersection could solve the problem by making capacity exceed counted demand. However, users of modern networks are likely to change routes as improvements are applied often rendering predictions of congestion solutions only dreams. Indeed, analysts now realize that there are few if any isolated intersections or streets in urban networks. Therefore improvements, even temporary changes like work zones, should be analyzed as a network problems. Network analysis tools however, require more traffic data than isolated intersection or arterial analysis tools. The most widely available source for such data is the demand estimation process that is developed and maintained by metropolitan planning organizations (MPO) in almost every urban area of the United States. The most popular demand estimation process is the "four step" model that was actually designed as a network analysis tool. In traditional applications, it was rarely used for operational analysis because the usual product was more representative of travel demands rather than actual traffic volumes. Recent innovations to four step modeling procedures, particularly Dynamic Traffic Assignment (DTA), have produced a very useful operational analysis tool. Additionally, the MPO modeling processes provide both current and forecasted future traffic estimates enabling both short term and long term analyses. The combination of DTA for predicting network traffic volumes and detailed micro-simulation procedures provides a true state-of-the-art tool set. However, researchers and practitioners need procedures for reducing DTA processing time and easy conversion of network characteristics to formats required for micro-simulation.

The Four Step Demand Estimation Model

There are four steps in the traditional transportation demand estimation process: 1) trip generation, 2) trip distribution, 3) mode choice, and 4) traffic assignment. There are two very different approaches to treating traffic conditions in the fourth step of the network transportation demand estimation process known as traffic assignment. The commonly used approach is called static traffic assignment (STA), while the newer and arguably more desirable approach is called dynamic traffic assignment (DTA). The

DTA technique provides very good approximations of link traffic volumes because it takes into account that traffic demand changes over time. Yet, DTA model processing time can literally require days for a single simulation of a large urban network. For planning purposes, long run times might not cause much concern, however, for operational purposes long run times are problematic. Usually, the simulation time of a network using DTA is around 20 times longer than with STA. A solution to the problem of long simulation times is the concept of using a subnetwork (Zhou et al., 2006; Chen et al., 2012; Gemar, 2013; Gemar et al., 2014; Bringardner, et al., 2014; Bringardner, 2015).

However, although DTA simulators, such as VISTA, use a simulation process (meso-simulation) to predict how travel times change with traffic volumes that type of simulation does not describe inter-vehicle activities. In order to have a comprehensive traffic analysis, an identical network can be examined using micro-simulation as a supplement to DTA to provide needed detail. However, this process of building or modifying a network in another simulator could prove to be time consuming and not very straightforward. For example, if DTA is the source of link volumes that feed the micro-simulation, such data must be gathered from DTA output and properly characterized for input to the micro-simulator. This effort as well as the effort required to describe the network typology for multiple scenarios for both DTA and micro-simulation is extensive. For larger networks the probability of more modifications to the traffic flow increases as does the level and complexity of the analysis effort. Thus, the ability to create the required input for micro-simulation directly from a DTA network specification would greatly reduce both network preparation and simulation time thereby enabling more complete analyses.

1.1 Background

The first three steps of the four-step transportation demand modeling process produces one or multiple OD matrices. The first step, Trip Generation, predicts the number of trip ends, called productions from and attractions to, to each zone centroid. The second step, Trip Distribution, connects the trip productions from each zone to the other zones developing the OD matrix according to the relative attractiveness and disutility associated with the other zones. Because the trip generation models generally produce person trips (not mode specific trips) the OD matrix, usually called the trip table,

includes trips by all modes of travel, including bus and private vehicle. The third step, Mode Choice, divides the person trip table into mode specific trip tables for each considered mode. The OD matrix for private vehicles can be assigned to the highway network and the matrix for public transportation can be assigned to that network (McNally, 2007).

The ability to predict link travel times as they change with traffic volume over time is a critical element of DTA. The most desirable technique for these predictions is a traffic simulation tool built into the DTA system. Three families of traffic simulators are available for application to DTA models, microscopic, mesoscopic, and macroscopic simulators. Yperman (2007) stated that a microscopic simulation tool is suitable only for small networks because the highly detailed predictions of vehicle reactions to other vehicles, traffic control, and geometry require very large execution times. A mesoscopic tool (like the cell transmission model in VISTA) produces a less detailed simulation by doing fewer calculations, so it can be feasibly used for medium sized networks. The macroscopic simulation family generally does not simulate individual vehicles, but characterizes only groups or platoons of vehicles and as such is generally not considered adequate for DTA application. The mesoscopic simulation family is generally most desirable for DTA, since it provides a sound basis for travel time estimation and tolerable run times.

However, since DTA involves many passes through a process that calculates minimum time paths, assigns traffic, and predicts travel times, the computational time is long if the network is large. For example, DTA application to the Dallas-Fort Worth urban network in the peak hours with VISTA requires almost two weeks of run time. If the network details and DTA techniques remain the same, then as computer technology becomes more advanced, the computational time will be less. However, DTA is not a static technology. DTA technology will change and include more network details, which will inevitably make computational times even longer.

Currently, certain pieces of information are estimated to simplify the evaluation of network problems. For example, origins and destinations are grouped into centroids (the center of the zone) rather than the actual locations. The collection of the demand data is usually by survey or by road-side equipment, such as a loop detector or radar. Thus, the collection of the detailed demand data requires additional amounts of time

and money. Furthermore, we currently simulate network operations only for selected times of the entire day.

As vehicles become more advanced, such as when connected vehicles are widely used, getting detailed traffic data will become more achievable. More details will lead to more all-inclusive and accurate OD matrices. If the network and the OD matrices become more comprehensive, the simulation period might change from only peak hours to the entire day. Therefore, as the DTA techniques advance, simulation run times may actually increase.

DTA simulation requires more complete network typology data compared to STA simulation. Since most MPO's use STA, their coded networks are only schematic, meaning that many links are simply not needed and therefore not included. Developing a minimum DTA detailed network requires a lot of time and money. Since most MPO's do not have coded comprehensive networks yet, DTA simulation is not a very feasible option. However, many network analysis questions involve changes that will not have significant impacts to most of the urban network. Such questions often include work zone traffic control issues and improvements to only a few links or intersections. If we can predict the size of the impacted area, and build a subnetwork that captures the significant changes, then the amount of required money and time can be reduced.

For such cases, Gemar (2013) analyzed the relationship between the capacity change of impacted links and the size of subnetwork that could capture all significant effects. This study defined the subnetwork size as acceptable when the OD matrices from the base subnetwork were similar to those from the modified subnetwork. In addition, this author designed a process for extracting a subnetwork from a regional network with ArcGIS and creating a subnetwork database with SSH, which is a network protocol that allows remote login and secure data transmission, and Visual Interactive System for Transportation Algorithms (VISTA), a mesoscopic DTA simulator. The reason for using ArcGIS at that time was that although TransCAD already had a coded network, it was a schematic network used for STA meaning that links were missing. Thus, CAMPO contracted the Center for Transportation Research (CTR) Network Modeling Center (NMC) at The University of Texas at Austin to complete the network so that it could be used for DTA. The NMC decided that the simplest way to complete the network for DTA purposes was with ArcGIS, which is a Geographic Information

System tool, because it provided a simpler way of editing, completing, and formatting the network for VISTA use in comparison to TransCAD. Later, Bringardner (2015) continued Gemar (2013) work by developing a Root Mean Square Error (RMSE) model, which estimates the optimal size of subnetwork given a user defined acceptable error.

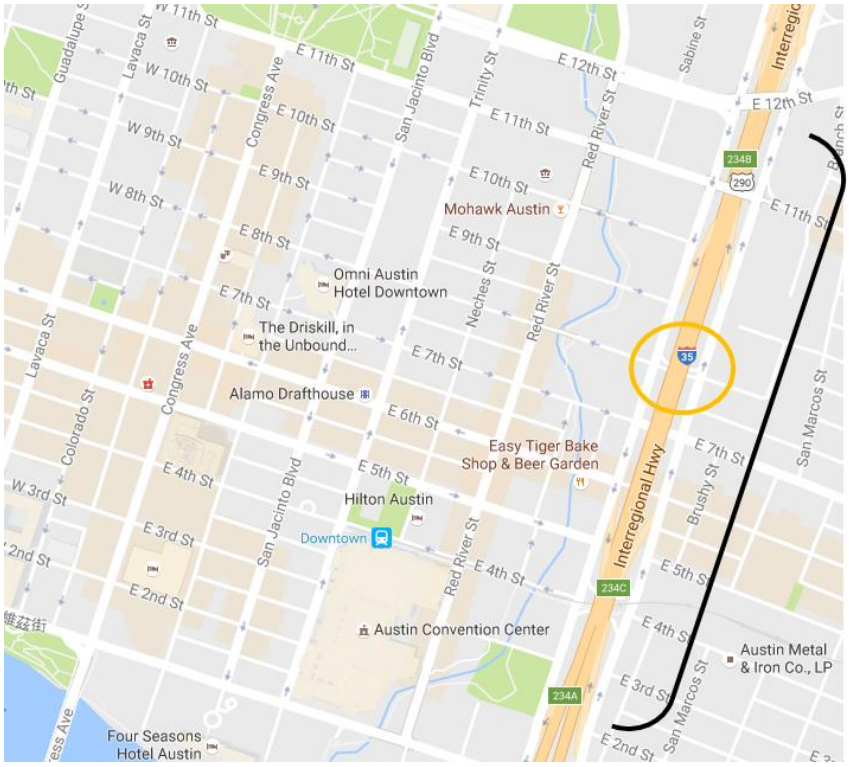
1.2 Motivation

Although the output of assigned traffic volumes from VISTA are a very good approximation of link volumes and the subnetwork extraction process can shorten the simulation run time, transportation engineers still encounter the following difficulties with the current methods:

1. The user must provide 3D network data for ArcGIS, otherwise elevated cross roads appear as false at-grade intersections.

Most traffic simulators see links and nodes as roads and intersections. Specifically, ArcGIS determines whether an intersection exists by determining if the links cross. However, a problem arises if the user imports network data that contains cross roads with different elevations as a 2D dataset rather than a 3D dataset. ArcGIS will see intersections where they do not exist, thus resulting in a larger subnetwork than expected. Figure 1 illustrates an example of a subnetwork that erroneously includes at-grade intersections with elevated freeway lanes. The yellow network is parameter size 5, the red and blue together is size 6, and the yellow, red, and blue together is size 7. In the golden circle, the two middle red links are the freeways, and the crossing yellow link included in parameter size 5 is an urban road. The freeway does not intersect with the urban road in reality, so these freeway links should not be recognized as intersecting with the urban road. However, since ArcGIS is not able to determine whether the crossed links actually represent at-grade intersections, these freeway links were erroneously identified as part of the parameter size 6 subnetwork. In ArcGIS this error carries over, which resulted in the links circled in blue being also included for parameter size 7 when they should not have been.

a)



b)

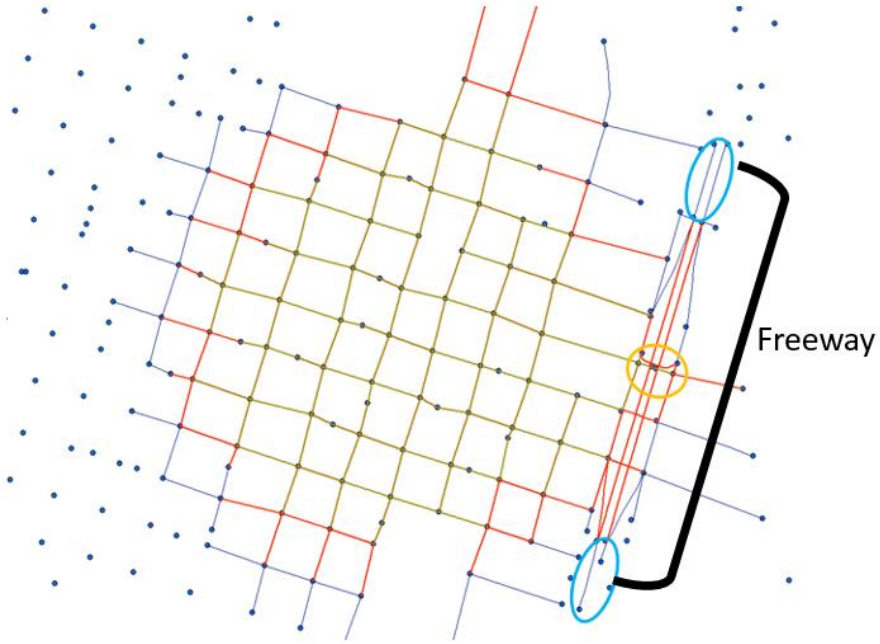


Figure 1 a) Map from Google Maps of the area and b) Example of an elevated crossroad misrepresented as an at-grade intersection in ArcGIS. In the golden circle, the two middle red links are the freeways, and the crossing yellow link included in parameter size 5 is an urban road, which in reality does not intersect with the freeway.

2. <u>The process of developing a subnetwork involves multiple manual steps that are prone to errors and may be complicated for some.</u>

Although the process from Gemar (2013) shortens the simulation time, it requires the user to install ArcGIS and SSH, and to modify JavaScript codes. Even though this author and CTR provide sample code, this could still be a complicated task. Therefore, this process might not be adopted widely without simplification.

3. The current subnetwork process lacks vehicle interaction data.

In order to shorten the simulation time, VISTA uses a mesoscopic simulator that does not perform detailed analysis of inter-vehicle actions, such as lane changing and or gap acceptance (Chiu et al., 2011). Such analyses can be developed through application of a microscopic simulator such as CORSIM or VISSIM. However, developing an identical VISTA subnetwork in a micro-simulation software is complicated and time consuming.

4. The current optimization model does not consider the simulation time.

Gemar (2013) developed a table which presents the relationship between the number of impacted links, the percentage of capacity reduction, and the subnetwork size to reduce the simulation time. Furthermore, since network error increases as subnetwork size and simulation time decrease, Bringardner (2015) developed RMSE models that find the optimal subnetwork size with the user defined acceptable error to reduce both simulation time and subnetwork size. However, the RMSE model procedure is not a traditional optimization process, because the user has to decide the acceptable error before running the models. In addition, both authors did not analyze the simulation time across different subnetwork sizes. While the subnetwork size increases a little bit, it may result in a much larger increase in the simulation time.

5. Micro-simulators do not consider vehicle path changing.

A micro-simulator input is link flow, and any change of geometry network (ex. adding an work zone) does not influence vehicle paths. Thus, practitioners usually build arterials or main streets in micro-simulators because they tend to have pretty stable traffic demand.

Most traffic tends to stay on the arterials for the shortest time path (more green time, higher capacity) when there is no interruption, so the highest impact from a work zone

is usually on arterial links. Figure 2 shows significant impacts from the work zone as blue links. Figure 2a shows no closure on the Guadalupe Street subnetwork and Figure 2b shows a 25% capacity reduction on Guadalupe Street. It is shown that only after a 75% capacity reduction (Figure 2d), the vehicle paths have significant change in the network. Otherwise, the work zone merely impacts the arterial's links.

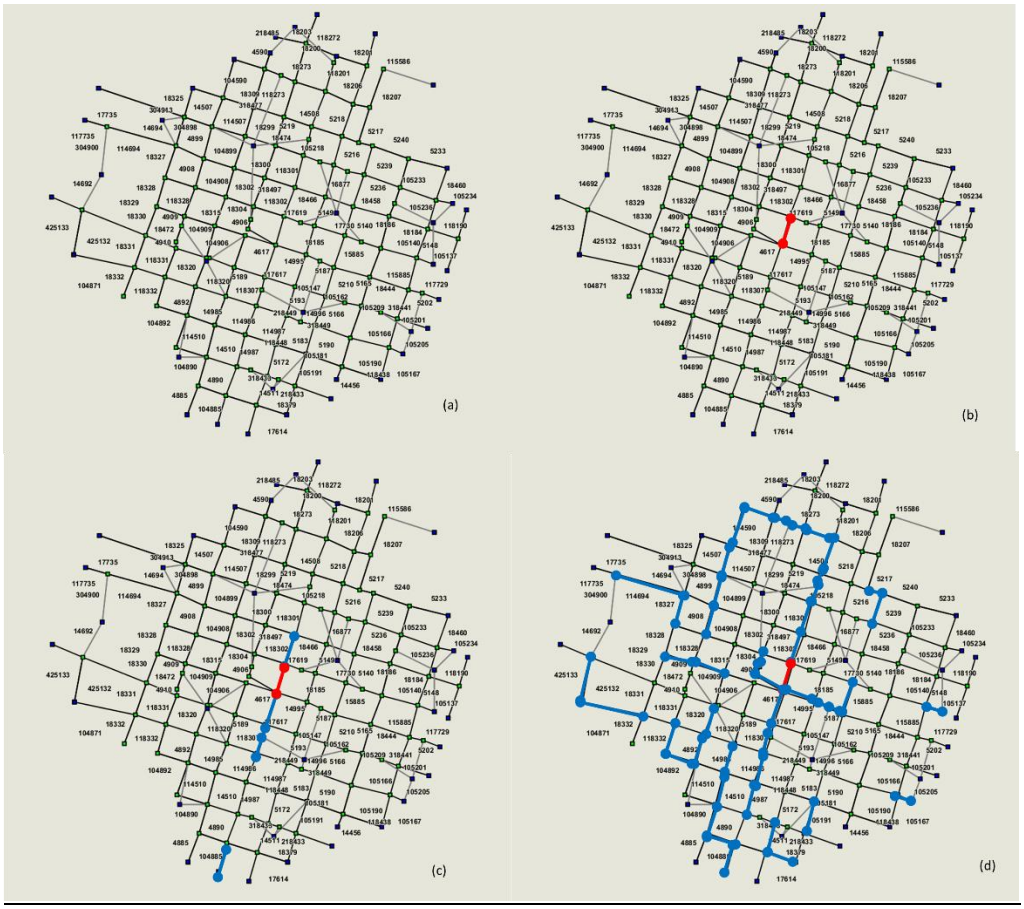


Figure 2 a) Base scenario using Guadalupe Street subnetwork and b) 25% capacity reduction on Guadalupe Street c) 50% capacity reduction on Guadalupe Street d) 75% capacity reduction on Guadalupe Street.

1.3 Research Objectives and Contribution

The primary objective of this research is to provide a robust system that will enable the automatic creation of an optimal size subnetwork from a regional VISTA network and an appropriate size network from the VISTA subnetwork for the CORSIM environment. There are two main components of the CORSIM network objective. First, develop a system that can automatically extract a micro-simulation-ready subnetwork

for CORSIM from VISTA. Second, develop a methodology that assesses the acceptability of the CORSIM subnetwork size.

In the first component, the system that this study aims to develop includes three tasks; first, remove the need for ArcGIS; second, automatically extract an optimal subnetwork from VISTA; third, automatically develop an identical subnetwork in CORSIM. To remove the need for ArcGIS, this study will develop a program to replace this software. The system considered the connection between links by node ID rather than as two-dimensional plots. If links that have the same node ID within the subnetwork boundary are treated as the next order link, then the problem of detecting elevated cross-roads will be solved. Users will only need to import the network data, which contains link ID's, including upstream and downstream node numbers, into the system. Users will also need to enter the location of capacity change(s) along with the nature of the change(s). Then the system will output the link ID numbers and node ID's into the subnetwork automatically.

To automatically extract the subnetwork from VISTA, the system contains two subsystems, one for extracting the geometry of the subnetwork and the other for creating a subnetwork database through SSH. This system only needs the network data, which contains link ID's including the upstream and downstream node ID's. The user chooses the impacted links. The first subsystem identifies the location of the impacted links (in streets or in freeways) and uses the adaptable model to determine the optimal size for the subnetwork. Then, the system runs the second subsystem for creating a database through SSH. Consequently, the user will get a complete VISTA subnetwork.

In order to develop an identical subnetwork in CORSIM, the study built the subnetwork in the micro-simulator automatically. The TRF file from CORSIM contains all the network information, including the link and node positions, the signal timing and turn movements at each intersection, as well as the entry node volumes. Also, since CORSIM provides a detailed user manual that describes how the TRF file is used, the information required to create the TRF file is readily available.

In addition, this study analyzed indexes that significantly influence the simulation time in the DTA simulator. These key indexes are an important reference for deciding the sub-area size, since the main reason for using a subnetwork is to reduce simulation time.

Thus, this study used key factors to build linear regression models for estimating simulation time. Using this estimated simulation time model, users define the optimal subnetwork size by integrating with the optimized RMSE models (Bringardener, 2014)).

This study estimated the CORSIM sub-arterial size whose link flows have significant difference between the base and impact scenarios using linear regression models. This study defined the significant difference as 95% statistical difference between the base and the impacted scenario. The random seeds determine the link flow in the base scenario, and the random seeds along with the artificial adjustment (i.e. work zone, etc.) determine the link flow in the impacted scenario. It is assumed that a link is significantly impacted by the adjustment when the link flow changes in the impact scenario more than the change in the base scenario. This study used statistical tests to define the maximum flow change in the base scenario; then developed two models, one to estimate maximum acceptable flow change in the base scenario, and the other to estimate the flow change using RMSE in the impact scenario. If the flow change for the impact scenario is larger than the maximum acceptable flow change for the base scenario, we can identify the link as having a significant impact from a change in the network.

Currently researchers and practitioners use very complex tools to analyze network problems. This research is going to provide a new tool for operational purposes that will allow users to apply very detailed, complex tools more easily to a variety of network issues. With this tool, researchers and practitioners will not have to build the sub-network manually in a DTA simulator and will be able to easily transfer that sub-network to a microscopic traffic simulator. In addition, researchers and practitioners can define the optimal subnetwork size not only by taking into account the acceptable error, but also the estimated simulation time. Finally, researchers and practitioners will be able to analyze traffic conditions on the impacted arterial, rather the entire sub-network in a microscopic traffic simulator, by using the sub-arterial size linear regression model.

1.4 Data

The data for this program came from the Network Modeling Center (NMC) at the Center for Transportation Research (CTR). The Downtown Austin dataset is from 2005

and includes all the streets (one-way, two-way), bus routes, signal timings, and OD matrices. The Downtown Austin area is west of I-35, east of North Lamar Blvd, north of Cesar Chavez Street, and south of 38th Street. Note that the I-35 freeway is coded in this data set. This program was developed using C++ and the subnetwork was developed in VISTA. In addition, the detailed process of developing a subnetwork database in VISTA came from Gemar and Bringardener's previous work. The RMSE models used were from Bringardner (2015), meaning that the subnetwork sizes defined in this study are the same and the resulting sub OD matrices are similar to the previous study. Using the models from Bringardner provided a comparison basis for making sure that the program works. The reason why the resulting sub OD matrix is similar rather than identical is that VISTA has random seeds, which affects vehicle rerouting in every simulation.

1.5 Overview

This dissertation will be divided into seven chapters. The first two chapters introduce the problem background and the unsolved problems of the current solutions. The next three chapters describe the proposed solutions. The final chapters constitute the desired goal of this study, including the implementation of the solution methods.

Chapter 1 Introduction This chapter provided an overview of the background, motivation, and problem statement.

Chapter 2 Literature Review This chapter will compare Static Traffic Assignment (STA) and Dynamic Traffic Assignment (DTA), introduce DTA simulators, VISTA and MatSIM, discuss the current knowledge associated with the creation of subnetworks and the use of DTA subnetwork analysis, introduce micro-simulators, CORSIM and VISSIM, and introduce the factors of the computer performance.

Chapter 3 Data This chapter describes the sub-area's configuration, including the location, capacity reduction, work zone sizes, and sub-area sizes.

Chapter 4 Simulation Time This chapter describes the indexes most affecting DTA simulation time, and uses these indexes to develop the regression models. Also, this chapter describes the optimal functions of the simulation time and size by following the Bringardener (2014) estimated sub-area size model.

Chapter 5 Linear Regressions for CORSIM sub-arterial size This chapter proposes models to identify the sub-network size in CORSIM and the links that have more than 95% statistical difference.

Chapter 6 DTA subarea developer This chapter describes the process of automatically extracting and developing a VISTA subnetwork.

Chapter 7 CORSIM TRANSFORMER This chapter proposes a way of developing a CORSIM network that is identical to a VISTA sub-network.

Chapter 8 Conclusion The final chapter summarizes the overall contributions and issues that should be addressed when applying the methodology in practice.

CHAPTER 2: LITERATURE REVIEW

The traditional transportation planning demand estimation model has been used for evaluating the impact of changing traffic conditions due to changes in demographics, network geometry, or land use (Chiu et al., 2011). In the transportation planning model, traffic assignment is critical for predicting link flows or path travel times when there are changes to the network. Tom V. Mathew [1] summarized some of the major aims of traffic assignments as follows:

1. To estimate link traffic flow and possibly intersection turn movements,

2. To estimate path travel time from origins to destinations for a given demand,

3. To aggregate network measures, such as total system travel time, and

4. To estimate and analyze the paths used between each origin to destination (O-D) pair.

2.1 Traffic Assignment

The traditional transportation planning demand estimation model has four steps: trip generation, trip distribution, mode choice, and traffic assignment. A forecast OD matrix usually called a trip table is obtained from the third step; then, traffic assignment assigns the trips to chosen network paths for which travel times are estimated using one of several alternative processes. The primary steps in traffic assignment, also called network loading are shown in Figure 3.

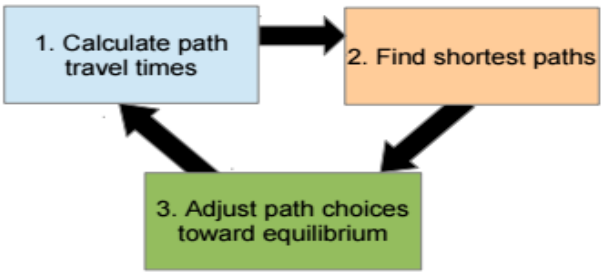


Figure 3 Primary steps of traffic assignment[2]

[1] Tom V. Mathew website: https://www.princeton.edu/~alaink/Orf467F08/TransportationNetworkDesign_Mathew.pdf

[2] Boyles (2016). Class notes, which is available at https://dl.dropboxusercontent.com/u/48099860/www/teaching/ce392d/class2.pdf

The steps include determining minimum route travel times for every origin to every destination based on free flow travel times. As traffic from the trip table is assigned to the routes with minimum time paths, travel times adjust to account for speed reductions caused by increasing numbers of trips. By some methods, new minimum time paths are calculated and more traffic is loaded on these paths followed by calculation of new travel times and new minimum time routes. During the iterative process, some algorithms may shift link flows from paths with longer travel times to paths with shorter travel times. The progress loop ends when the assignment finishes allocating the specified trip interchanges. There are different ways of allocating the trip interchanges, with the simplest and quickest being the all-or-nothing method, followed by more complex methods like system optimal and user equilibrium techniques.

All-or-Nothing Assignment

An all-or-nothing assignment assigns all travel demand between any origin and all destination nodes to single minimum cost paths where path cost or time is based on fixed user specified travel speeds. Other paths with larger cost are not assigned any traffic volume. The technique is simple and fast; it estimates drivers' desired paths without considering congestion or the capacity constraints. However, the traffic flow from this technique is unrealistic because congestion and capacity are ignored (Meyer, M. D., and Miller, E. J, 2001). This assignment can be used in uncongested networks (Mathew and Krishna Rao, 2006). An all-or-nothing assignment is usually the first step of system optimal or user equilibrium assignments.

System Optimization (SO) Assignment

The SO assignment method aims to assign traffic demands to paths that produce the minimum total system cost and follows Wardrop's second principle, which states that *drivers cooperate with one another in order to minimize total system travel time* (Mathew, T. V. and Krishna Rao, K. V, 2006). This assignment can be thought of as a model in which congestion is minimized when drivers are told which routes to use. Obviously, this is not a behaviorally realistic model, but it can be useful when trying to manage traffic to minimize travel costs and therefore achieve an optimum social equilibrium. The process of this assignment is: first, define the starting link cost (ex. free flow travel time); second, assign the traffic by the all-or-nothing method; third,

search all the alternative paths for the same OD to find the path that has the minimum marginal cost; fourth, assign the traffic flow to the path with minimum cost; fifth, repeat the third and fourth steps for all the ODs; sixth, calculate all the marginal path costs; and seventh, repeat steps four through six until all paths with the same ODs have the same marginal cost.

User Equilibrium (UE) Assignment

The user equilibrium assignment is based on Wardrop's first principle, which states that *no driver can unilaterally reduce his/her travel costs by shifting to another route*. If it is assumed that drivers have perfect knowledge about travel costs on a network and choose the best route according to Wardrop's first principle, then this behavioral assumption leads to deterministic user equilibrium. UE means all the route travel times for paths serving the same OD are similar or identical. Network travel time from user equilibrium might be larger and more realistic than from system optimization because the drivers will optimize their own path travel times in UE. In SO, alternative paths for any OD pair might have different travel times. However in UE, paths with the same OD should have the same travel time even though the system travel time may be greater than for SO. The process is similar to system optimization, except for the third, sixth, and seventh steps. The third step in user equilibrium assignment is to search all the paths with the same OD pair and find the minimum cost path. The sixth step is to calculate all path costs, and then repeat steps four through six until all the possible paths for each OD pair have the same path cost.

There are two families of approaches commonly used for traffic assignment: static traffic assignment (STA) and dynamic traffic assignment (DTA).

2.1.1 STATIC TRAFFIC ASSIGNMENT (STA)

STA is usually used to simulate a relatively long time-of-day period, such as a multi-hour peak period, and it assumes that the traffic flow does not change over time. In STA, the entering flow to a link equals the leaving flow during a specific time period. The link travel time is described by a link performance function that expresses the average or steady-state travel time on a link as a function of the traffic volume on the link (Chiu et al., 2011). According to the link performance function, link travel time increases significantly when traffic volume reaches or exceeds link capacity. Equation

(1) shows one of the commonly used link performance functions from the Highway Capacity Manual (HCM).

$$t_{ij}(x_{ij}) = t_{ij0}(1 + a(x_{ij}/C)^b) \,, \quad (1)$$

where t_{ij} is travel time from link i to j, x_{ij} is traffic flow from link i to j, t_{ij0} is free flow travel time from link i to j, a and b are Bureau of Public Roads (BPR) parameters (Typically a = 0.15 and b= 0.84), and C is link capacity.

The iterative mechanics of STA assignment processes vary considerably, however, an example of one is described as follows:

When the link travel times are updated, the shortest path travel time is chosen, and the path flow shifts to the path with shortest travel time. Equation (2) displays how the path flow shifts for each iteration.

$$x' \leftarrow (1 - \lambda)x + \lambda x^* \quad , \quad (2)$$

where x is the current path flow, $x^*$is the all or nothing path flow and λ can be defined by the Method of Successive Averages (MSA)[3] or the Frank-Wolfe (FW) [4] algorithm. In MSA, $\lambda = \frac{1}{1+I}$, and I is the number of iterations. This implies that x is getting closer to the optimal x as the number of iterations increases. On the other hand, Frank-Wolfe chooses λ, which is a smarter way to loop through Equation (3) – Equation (6). The goal is to choose λ so as to produce an x that is closest to equilibrium.

$$f = \sum t_{ij}(x'_{ij}) = \sum t_{ij}((1 - \lambda)x + \lambda x^*), \quad (3)$$

$$f' = \sum t_{ij}((1 - \lambda)x + \lambda x^*) * (x^* - x), \quad (4)$$

$$f' = \sum t_{ij}((1 - \lambda)x + \lambda x^*) * (x^* - x)^2, \quad (5)$$

$$\lambda \leftarrow \lambda - \frac{f'}{f''}, \quad (6)$$

[3]Boyles (2016). Transportation Network Analysis class notes, which is available at https://dl.dropboxusercontent.com/u/48099860/www/teaching/ce392c/class8.pdf

[4]Boyles (2016). Transportation Network Analysis class notes, which is available at https://dl.dropboxusercontent.com/u/48099860/www/teaching/ce392c/class10.pdf

where $t_{ij}(x'_{ij})$ is the path travel time and f is the sum of path travel times, f' is the derivative of f, and f'' is second derivative of f. After the initial λ is set between 0 and 1, the process from Equation (3) to Equation (6) repeats until the difference between successive λ's is less than the tolerance error. The optimal λ is achieved when all the path travel times for any given OD pair are identical.

2.1.2 DYNAMIC TRAFFIC ASSIGNMENT (DTA)

Dynamic Traffic Assignment (DTA) accounts for the time-dependence in traffic flows and conditions by using multiple OD matrices. DTA aims to reflect the reality that traffic networks are generally not in a steady-state. Hence, the path travel time is affected by the departure time.

The user equilibrium in DTA, called Dynamic User Equilibrium (DUE), is achieved when the vehicles with the same origin, destination, and departure time, have identical path travel times. Also, DTA does not assume that inflow equals outflow. As the inflow of a link becomes larger than outflow of the same link in the same period, its density increases and speed decreases, consequently, travel time and delay increase, leading to congestion (Chiu et al., 2011). Since link travel time is affected by the downstream link, and the path travel time is the sum of the experienced link travel times, the only way to identify whether user equilibrium is achieved is through simulation.

Link and Node Simulation Models

A simulation can run as either a node or a link model. A node model defines how much flow can move from the upstream link to the downstream link according to the sending flow and receiving flow from the upstream and downstream links. A link model defines the sending flow and receiving flow from the link according to the link density, free flow speed, and link length. Receiving flow is the number of vehicles that will enter the link if there is no obstruction from the upstream link, and sending flow is the number of vehicles that will leave the link if there is no obstruction from the downstream link[5].

There are four types of link models, Point Queue, Spatial Queue, Link Transmission Model (LTM), and Cell Transmission Model (CTM). Point Queue, Spatial Queue, and

5 Boyles (2016). Dynamic Traffic Assignment class notes, which is available at https://dl.dropboxusercontent.com/u/48099860/www/teaching/ce392d/class9.pdf

LTM models are intended for macroscopic simulations, which present only inflow and outflow of each link. Thus, how the vehicles move within the links is not shown. CTM is a mesoscopic simulation model, which splits a single link into several cells. CTM shows the number of vehicles in cells, therefore the traffic flow or density change at each time interval in each cell is known. However, CTM does not show how the vehicles interact with each other.

Point Queue, the earliest link model, was presented in the 1970s. According to Ban (2012), the Point Queue model stores link volumes as vertical queues at the end of the link, which implies that a link is a point, and that vehicles stack on top of this point. Therefore, the queue lengths and dissipation rates at the links determine delay and travel time. This is a macroscopic model that results in simple and concise mathematical formulations. However, Point Queue cannot capture flow propagation at the link level or congestion spillback over consecutive links because it assumes that vehicles stack on a point instead of queue on a link. Equation (7) and (8) present the sending and receiving flow from the Point Queue model[6].

$$S(t) = \min\left(N^{\uparrow}\left(t - {}^{L}/_{u_f} + \Delta t\right) - N^{\downarrow}(t), q_{max}\Delta t\right), \quad (7)$$

$$R(t) = q_{max}\Delta t\,, \quad (8)$$

where t is time, $S(t)$ is sending flow at t , $R(t)$ is receiving flow at t, $N^{\uparrow}$ is the actual traffic flow at the upstream end of the link, $N^{\downarrow}$ is the actual traffic flow at the downstream end of the link, L is the link length, u_f is free flow speed, q_{max} is maximum traffic flow, and Δt is the time interval. The sending flow is the current number of vehicles on the link or the maximum traffic flow.

The Spatial Queue link model adds the restriction that links have a maximum number of occupying vehicles, whereas Point Queue does not limit the number of queued vehicles. The entering flow is reduced when the link density reaches the maximum link density. Although the Spatial Queue model limits the number of vehicles in a link, the model does not consider the time required for congestion spillback. Equation (9) and

[6] Boyles (2016). Dynamic Traffic Assignment class notes, which is available at https://dl.dropboxusercontent.com/u/48099860/www/teaching/ce392d/class11.pdf

(10) present the sending and receiving flow from the Spatial Queue model, respectively[6].

$$S(t) = \min\left(N^{\uparrow}\left(t - {}^{L}/_{u_f} + \Delta t\right) - N^{\downarrow}(t), q_{max}\Delta t\right), \quad (9)$$

$$R(t) = \min(N^{\downarrow}(t) + k_j L - N^{\uparrow}(t), q_{max}\Delta t), \quad (10)$$

where t is time, $S(t)$ is sending flow at t , $R(t)$ is receiving flow at t, $N^{\uparrow}$ is the actual traffic flow at the upstream end of the link, $N^{\downarrow}$ is the actual traffic flow at the downstream end of the link, L is the link length, k_j is maximum density of the link, q_{max} is maximum traffic flow, and Δt is the time interval. Unlike the Queue Point model, the receiving flow is limited when the link density reaches the maximum.

LTM adds in reaction time to avoid the Spatial Queue model problem of unknown congestion spillback time. Equation (11) and (12) display the receiving and sending flow from LTM,

$$S(t) = \min(N^{\uparrow}\left(t - {}^{L}/_{u_f} + \Delta t\right) - N^{\downarrow}(t), q_{max}\Delta t), \quad (11)$$

$$R(t) = \min(N^{\downarrow}\left(t - {}^{L}/_{w} + \Delta t\right) + k_j L - N^{\uparrow}(t), q_{max}\Delta t), \quad (12)$$

where w is vehicle speed in a congested situation. Equation (12) includes $-{}^{L}/_{w} + \Delta t$, accounting for the upstream spillback time.

Lastly, CTM is a mesoscopic simulation model that functions similarly to a node model in that the model splits links into different cells based on free-flow speeds. The traffic flow between adjacent cells is determined by the link capacity, the maximum available downstream cell entering flow, and the entering flow from the upstream cell. Since the number of vehicles in each cell is known, identification of congestion development is possible[7].

2.1.3 STA VERSUS DTA

STA assumes that traffic demand is steady and that the departure time does not affect the path travel time. Therefore, the path travel time is the sum of the link travel times

[7] Boyles (2016). Dynamic Traffic Assignment class notes, which is available at https://dl.dropboxusercontent.com/u/48099860/www/teaching/ce392d/class8.pdf

(also called instantaneous travel time) calculated by the link performance function. The variable in the link performance function is the link flow. Although STA is fast and simple, this assignment is unable to represent real conditions for three reasons:

First, real traffic conditions are dynamic, not static,

Second, STA does not generally restrict link flow to be less than the capacity, and

Third, STA equalizes the inflow and outflow.

In DTA, the travel demand for OD pairs changes over time like real network demands and link travel times are affected by downstream links. Therefore, path travel times are good approximations of the actual experienced times as they are affected by the departure times. Time based OD matrices must be characterized by field data that would likely not be required for an STA assignment. Of course, the field data must be an accurate representation of reality or the traffic volume produced from DTA will be inaccurate. Nevertheless, DTA requires multiple OD matrices, one for each time increment to be examined, or a function that describes how demand changes with time, which requires a considerably larger amount of computational power.

In reality, congestion occurs when outflow from the upstream link is larger than the maximum inflow for the downstream link. Vehicles reaching the end of the upstream link slow down and wait until there is adequate space to move on. The slow moving vehicles at the end of the upstream link affect later arriving vehicles, and so on. Congestion spills back to the upstream link when the queue length is greater than the link length. Hence, the link travel time is affected by the downstream link capacity or maximum available inflow rather than its own capacity.

Lastly, link inflow and outflow are not identical in real traffic conditions. Instead, what we call traffic flow in STA is more similar to traffic demand rather than actual traffic volumes. Since STA works with a simple "schematic" network, traffic engineers can use a less detailed network that does not contain all actual links to find traffic demands. Conversely, DTA produces link flows that are much closer to actual traffic volumes because the process by which congestion develops and assumptions of traffic conditions are similar to reality.

2.2 Subnetwork Analysis

A subnetwork is a portion of an entire network, which could be as small as several blocks or almost as large as the entire network. In Figure 4, a work zone/impacted link that is the subject of the analysis effort is shown in yellow, the subnetwork is shown in red, and the entire network is made up of all the red and green links. Subnetworks are often chosen to enable use of detailed network analysis methods for operational questions about very large networks.

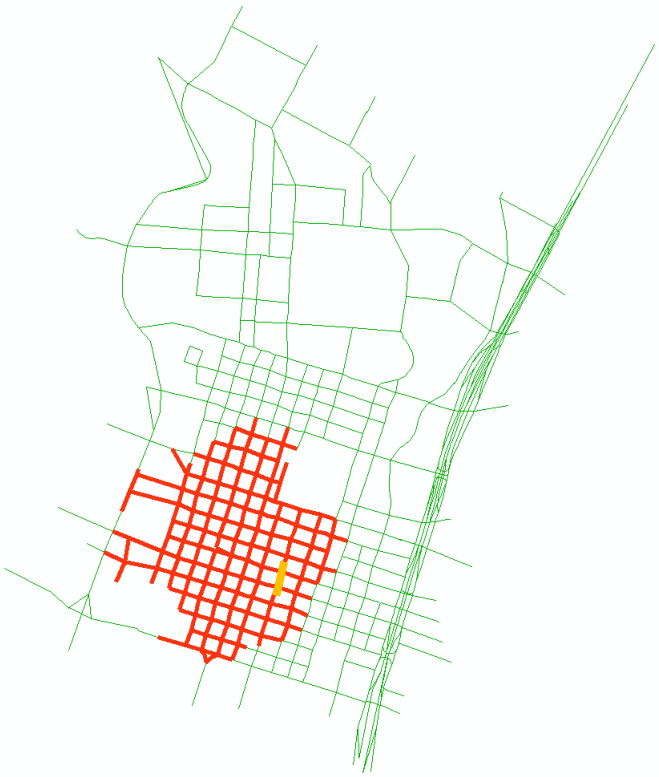


Figure 4 Example of subnetwork in the entire network

Gemar (2013) compared the computational time and effort for the full Dallas-Fort Worth network to a Dallas subnetwork and found the required computational time and effort are reduced. The respective computation time and effort reductions were approximately 1022 hours and 57 GB in the full Dallas-Fort Worth network to around 73 hours and 10 GB in the Dallas subnetwork because of the reduced number of vehicles, links, and nodes (Table 1 and Table 2).

Table 1 Example of the reduction of simulation time in Dallas-Fort Worth network (Gemar, 2013)

Network	DTA Analysis Processes	Path Generation Iterations	DUE Iterations	Time to Complete Analysis (hours)	Calendar Days to Complete
Full DFW Network	32	8	36	1,022.3	47
Dallas Subnetwork - Base	15	20	40	72.2	5
Dallas Subnetwork - Impact	12	20	40	73.8	5

Table 2 Example of the reduction of file space required for Dallas-Fort Worth network (Gemar, 2013)

Network	Number of Links	Number of Nodes	Number of Vehicles	Final Gap Measure	Total File Space Required
Full DFW Network	71,721	31,364	2,512,462	13.00	57.24 GB
Dallas Subnetwork – Base	16,434	7,408	858,613	7.15	9.73 GB
Dallas Subnetwork - Impact	16,414	7,394	858,613	7.80	10.17 GB

Overall, the required time and space were reduced even though the number of iterations in the subnetwork was higher than in the entire network. With the notable exception of area-wide long-range demand estimation projects, most network questions about traffic involve changes to only a few links of a very large network. Analysis of such questions does not necessarily require inclusion of the entire urban network, but only a small part of the total network. Since the subnetwork shortens the computational time and effort, it is possible that a subnetwork could provide perfectly acceptable answers to operational questions without including the entire network. However, the accuracy of the performance of the subnetwork is an issue that requires careful consideration.

The required size of the subnetwork is related to the magnitude of property changes of the impacted link(s) (e.g. the capacity reduction). The changed impacted links cause rerouting of some vehicles and the subnetwork size must be adequate to include the possible rerouting options. But on the other hand, the full network size has a much

longer simulation time, which can make the subnetwork a more attractive option even if some rerouting is not included.

Chen et al. (2012) called the links where changes are proposed and the connected links upon which traffic volumes change significantly the *impacted* area, and defined that area with a size parameter. The subnetwork with a size parameter of one includes an impacted link and all the links adjacent to the impacted link. The subnetwork with a size parameter of two includes the network with a size parameter of one and the links adjacent to this network **(**Figure 5**)**. The links of the next size parameter are the connected links to the links in the previous size parameter.

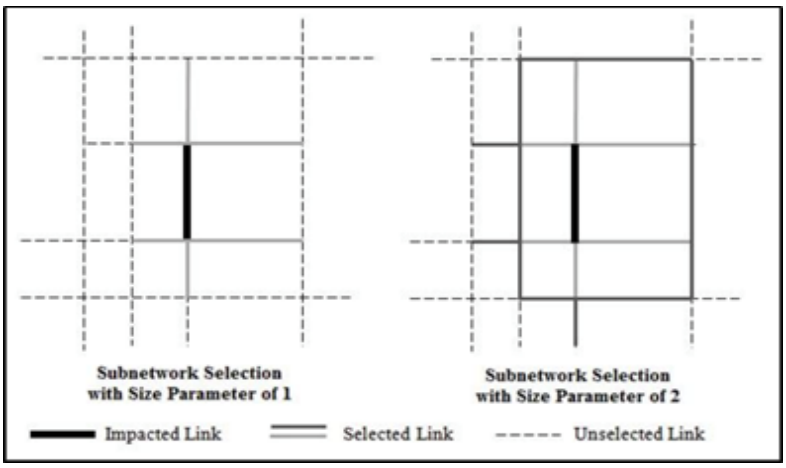


Figure 5 Visualization of the Subnetwork Selection with size parameter Process (from Gemar, 2013)

2.2.1 SUB NETWORK AND SUB OD MATRIX DEVELOPMENT

Gemar (2013) developed a process for identifying a subnetwork and a sub OD matrix that requires ArcGIS, which is a geographic information system (GIS) platform, SSH, which is a network protocol that allows remote login and secure data transmission, and the Visual Interactive System for Transportation Algorithms (VISTA), a mesoscopic DTA simulator. This process includes five steps:

The first step is to determine the subnetwork size in the whole network, which was defined by order number as referenced in Chen et al. (2012).

The second step is to isolate the subnetwork elements from the whole network with ArcGIS, including links, nodes, connectors, and centroids. The process includes,

1. Turn on the link layer and select the impacted links.
2. Select *Order Selection* and type the size order in. ArcGIS captures the links in this order. Then, isolate these links to the subnetwork link layer.
3. Turn on the node layer.
4. Select the subnetwork nodes by using the option "*touch the boundary of the source layer feature*", choose the subnetwork links as source layer, and add the selection of within distance of the source layer (20 feet). The subnetwork nodes would be captured. Then, isolate the captured nodes to the subnetwork node layer.
5. Turn on the connector layer.
6. Select the subnetwork connectors by choosing the option "*touch the boundary of the source layer feature*", and choose the subnetwork nodes layer as source layer. Then, isolate the captured connectors to the subnetwork connector layer.
7. Turn on the centroid layer.
8. Select the subnetwork centroids by using the option "*touch the boundary of the source layer feature*", choose the subnetwork connector as source layer, and add the selection of within distance of the source layer (25 feet).
9. Save subnetwork links and connectors' ID into link data, and save subnetwork nodes and centroids' ID into node data

Third, duplication of the complete network in the VISTA database with SSH;

Fourth, import the link data and node data into the VISTA database with SSH;

Fifth, replace the complete network link and node data with the subnetwork in the copied network database with SSH; and

Fifth, identify the routing data with SSH to establish the sub OD matrix. The code for developing a sub OD matrix was received from Network Modeling Center (NMC) at the Center for Transportation Research (CTR) at The University of Texas at Austin (UT Austin). The location of the whole network and the subnetwork must be added in before executing the SSH code.

2.2.2 SUBNETWORK EVALUATION

An induced origin-destination matrix, developed by Larsson et al. (2001), is the best way to create a sub OD matrix. The vehicle paths, i.e. the series of links that the vehicles have crossed, from the regional model identify the origins and destinations of the subnetwork. Traffic movements in the sub OD matrix exclude vehicles that do not cross the subnetwork. Also, the time from origin to destination in the sub OD matrix comes from the subnetwork arrival and subnetwork departure times. Figure 6 is a schematic of the departure and arrival times in the subnetwork. The departure and arrival times for the vehicles crossing the full network are 6:30 pm and 7:00 pm, however, for the subnetwork the times are 6:37 pm and 6:55 pm, respectively.

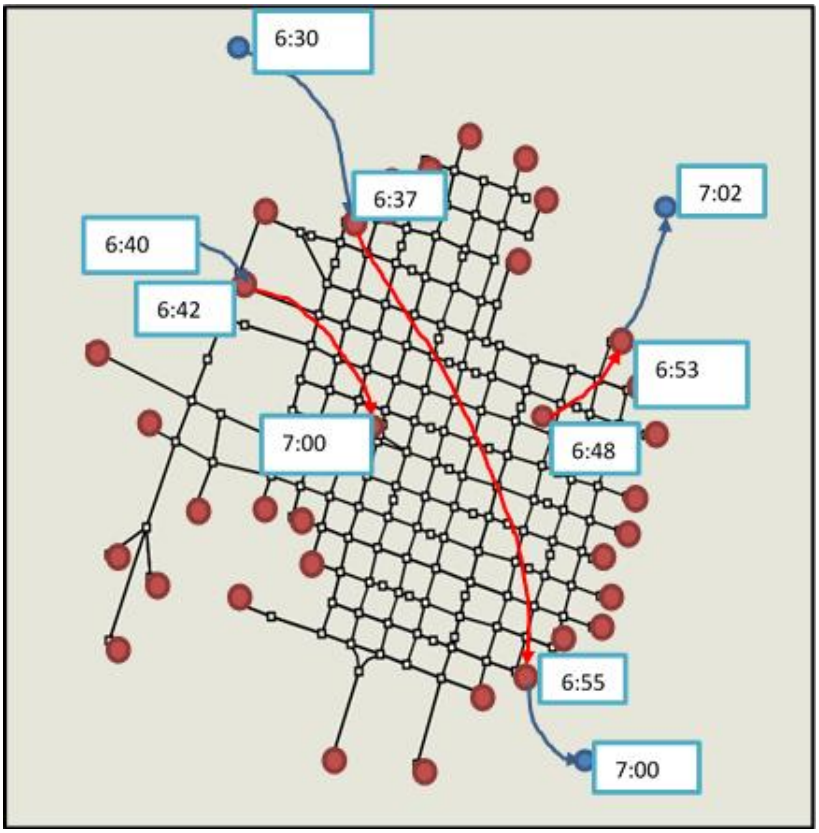


Figure 6 Schematic of the departure time and arrival time in the subnetwork

Bringardner et al. (2014) referred to the size parameter from Chen et al. (2012) and investigated measures that could evaluate the quality of the subnetwork size. Root Mean Square Error (RMSE), Mean Censored Absolute Percent Error (MCAPE), and SSIM (Structural Similarity) indices were examined by comparing the base sub OD matrix to the impacted sub OD matrix. The authors found that RMSE was the most effective

measure because RMSE has the ability to show a significant error difference between the impacted and base subnetwork OD matrices and other measures could not. RMSE shows the difference between a single simulation and the average simulation. Both studies used RMSE to evaluate subnetwork validity and to study what size of subnetwork can sufficiently replace the entire network. Equation (13) shows the calculation of RMSE,

$$\mathrm{RMSE} = \left(\frac{1}{n} * \sum(\overline{d_{ijt}} - d_{ijt})^2\right)^{0.5}, \qquad (13)$$

where n is the number of simulations, i is origin, j is destination, t is time period, and $\bar{d}$ is the base scenario average demand, d is the subnetwork demand extracted from a simulation (Bringardener et al., 2014).

A sub OD matrix might change when the subnetwork elements change (e.g. link capacity reduced) even though the size of the subnetwork is fixed. Zhou et al. (2006) updated the boundary demand with the result from a push-based STA to shorten the computational time. These authors used a logit model to remove the trips with path travel time in the subarea larger than in the rest of the subnetwork. To simplify the problem, they did not analyze the trips that originally traveled from the subarea to the rest of the network, and from the rest of the network to the subarea.

Gemar et al. (2014) adjusted the OD matrix of a subnetwork based on a logit model, which is a type of regression model with a two-level categorical dependent variable. The utility function of this logit model is travel time. The travel time is split into two types to indicate whether a vehicle whose origin and destination are outside the subnetwork is entering this subnetwork or not. One is external travel time and the other is internal travel time. The external travel time is assumed to remain constant even though traffic conditions change inside the subnetwork. The number of vehicles whose travel time going through the subnetwork is shorter than not going through the subnetwork must be counted in the sub OD matrix.

In order to simplify the subnetwork OD matrix, Gemar et al. (2014) grouped zone centroids external to the subnetwork into megacentroids. Three grouping methods, namely, grouping by region, by maximum demand, and by demand proportion were investigated by simulating the downtown Austin subnetwork and calculating Root

Mean Square Error (RMSE). Grouping by demand proportion proved to give the most realistic traffic conditions.

The method of adjusting the OD matrix can work in any size subnetwork, however, the sub OD matrix must be modified whenever subnetwork size or subnetwork traffic conditions change. Thus, it is desirable to define an optimal subnetwork size to avoid repeated adjustment of the sub OD matrix.

To avoid changes in the sub OD matrix and maintain the accuracy and efficiency of the subnetwork, Gemar (2013) studied the relationship between capacity reduction of the impacted links, the number of impacted links, and the impacted area (Figure 7). As the number of impacted links or the capacity reduction of the impacted links increases, the size of the subnetwork increases until the minimum acceptable subnetwork size equals the size of the entire network.

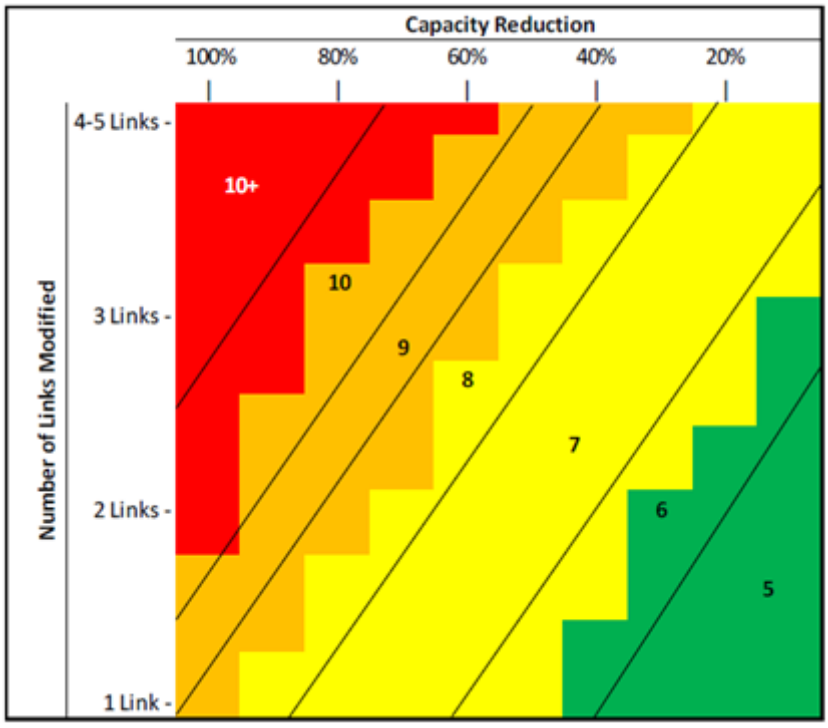


Figure 7 Recommended Subnetwork Sizes (from Gemar, 2013)

In addition, Chen et al. (2012) suggested that the range of the network size parameter be from 3 to 10 for subnetworks of any network. If the size parameter is less than 3, then most critical links cannot totally be contained in the subnetwork. Additionally, if

the size parameter is larger than 10, then the computational time savings are not significant.

In some cases, reduction of computational time might be more important than the accuracy of the subnetwork results. For example, in the case of a very short deadline that precludes a full run, a less accurate subnetwork result may be better than no result. Thus, it is important to study the balance between the accuracy of the subnetwork and the reduced simulation time. Bringardner (2015) referred to Gemar's (2013) process of developing a subnetwork database and used the results to develop a model that predicts RMSE as a function of subnetwork size. The user must provide the acceptable error so that the model can suggest the optimal subnetwork size. Equation (14) and Equation (15) correspond to the RMSE in the base and modified subnetwork, respectively.

$$\text{Base}_{\text{RMSE}} = -0.115 * s - 0.018 * c_l/1000 - 15.487 * \frac{v}{c} + 6.004, \qquad (14)$$

$$\text{Impacted}_{\text{RMSE}} = 0.430 * s + 0.135 * r_c + 0.719 * n - 0.117 * \frac{c}{1000} - 9.610 * \frac{v}{c} - 0.016 * s * r_c, \qquad (15)$$

where s is subnetwork size, c_l is link capacity, v is volume from the impacted link and c is capacity from the impacted link, r_c is percent of capacity reduced, and n is number of impacted links.

From these equations, capacity is a link property, number of impacted links and percent capacity reduced are predetermined, and volume to capacity ratio comes from the results of the entire network. Increasing the subnetwork size results in the decrease of RMSE in both equations. The rate of decrease in Equation (15) is faster than in Equation (14). When both RMSE values are identical, the size parameter is optimal and considered to have acceptable error.

RMSE models have been examined in downtown Austin and downtown Dallas. The limitation of these models is that they were derived using results from only one work zone per subnetwork. The longest work zone examined had three connected links and equal capacity reduction ratios. Application of the method to a network with several disconnected work zones, such as a developing city, could be problematic unless the analyst handles each work zone as a separate case.

2.3 DTA program: VISTA and MatSIM

2.3.1 VISTA

The Network Modeling Center (NMC) of the Center for Transportation Research (CTR) at the University of Texas at Austin offers a DTA analysis tool using the web-based program, Visual Interactive System for Transportation Algorithms (VISTA). VISTA is a mesoscopic simulation tool based on the Cell-Based Transmission Model (CTM). For the fourth step of the transportation planning model, VISTA uses the OD matrices produced by the three previous steps (trip generation, trip distribution, and mode split) to assign vehicles to the network.

Figure 8 shows the downtown Austin network in VISTA. The red and green dots are centroids and nodes, respectively, and the black and blue lines are links and connectors. Usually, a road/street is treated as a link, and the number of lanes, speed limit, or capacity are defined as link properties. Nodes are the junctions (or intersections) between links, and if an intersection is signalized, then the signal timing is defined in the node properties. Additionally, a centroid is a type of node that reduces complexity and simplifies the network by representing all actual trip origins and destinations in a zone. Thus, centroids are virtual and all trips start from and end at centroids. Finally, centroid connectors are virtual links that connect each centroid to the network. Since centroid connectors are virtual, they do not limit the number of vehicles and speed, and have zero travel time. As the network is loaded, trips are assigned to minimum time network paths from origin centroids to destination centroids, and travel times are estimated using the CTM.

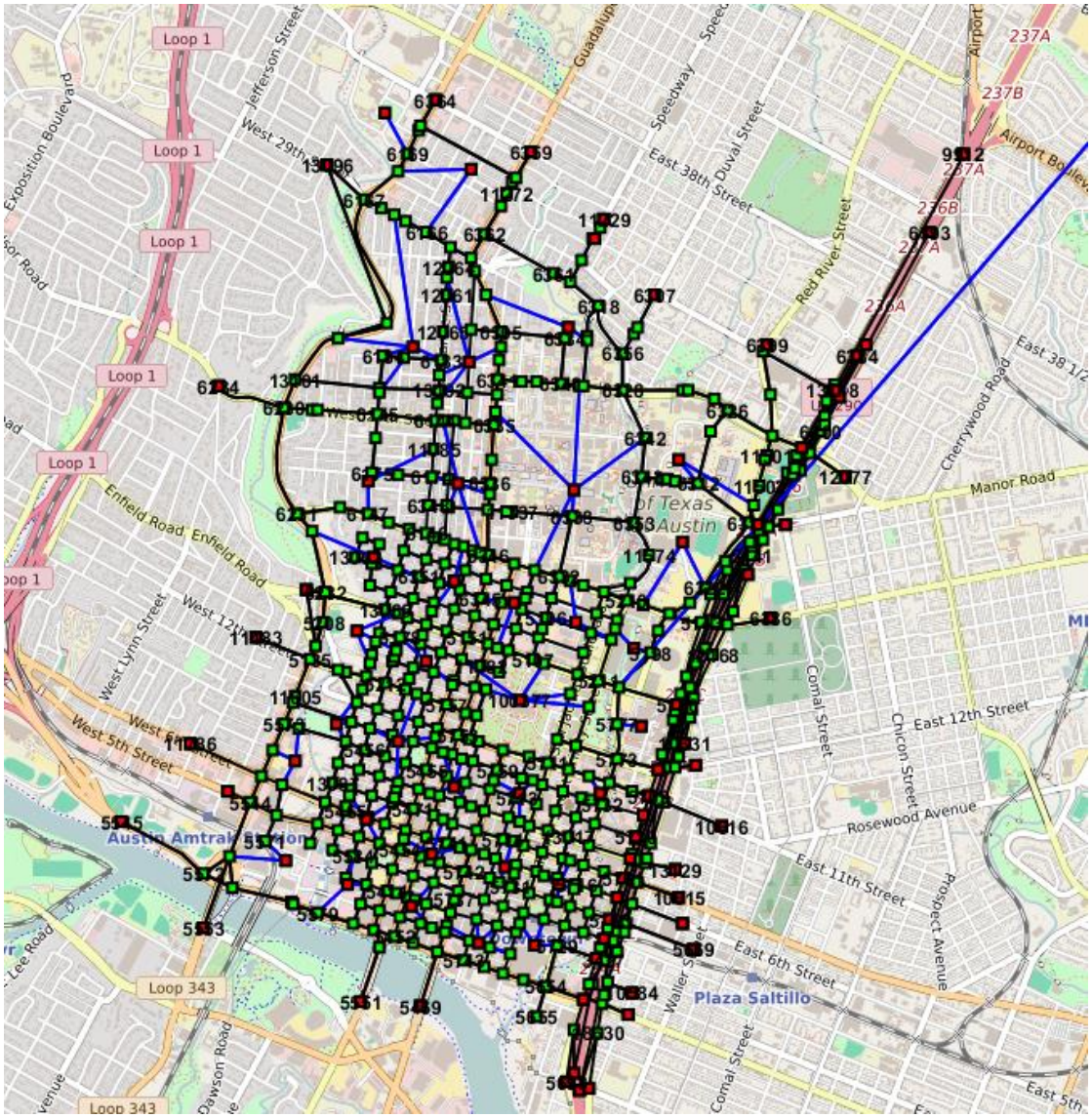


Figure 8 The Downtown Austin network in VISTA

2.3.2 MATSIM (MULTI-AGENT TRANSPORTATION SIMULATION)

MATSIM is an open-source large-scale agent-based dynamic traffic assignment simulator. Unlike the traditional transportation planning model, this agent-based model develops demand data by each agent's (person's) trip purpose, such as home-based work, home-based shop, or home-based school. Then, MatSIM tracks the agents throughout the entire day from home to work, to shop, or to leisure until the person arrives back home.

Figure 9 is an example of demand data in MatSIM. Each agent has a person id and multi-trip purposes, such as "h" means home and "w" means work. Moreover, "x" and "y" are the positions of the trip purpose, and the link is the departure or arrival link. If

the departure or arrival positions are available but the link is unknown, MatSIM can automatically find the closet link to the position. Although MatSIM is an agent-based simulator, MatSIM can still work with OD matrices used in the traditional planning model. Moekel (2016)[8] currently aims to develop OD matrices for MatSIM from a land use model.

```
<person id="1">
    <plan type="car">
    <act type="h" x="4438874.883070" y="5418430.612870" end_time="6:0:0"/>
    <leg mode="car">
    </leg>
    <act type="w" x="4436874.883070" y="5419400.449010"  />
    </plan>
</person>

%The person's id
%The person's transport tool is car
%Activity type is home to work
%Depart location is at (4438874.883070, 5418430.612870)
%Depart time is at 6 am.
%Arrive location is (4436874.883070, 5419400.449010)
```

Figure 9 MatSIM demand sample code and comments

Two critical data are necessary in MatSIM: the network and the demand data. Demand data were introduced in the previous paragraph. Network data can be downloaded from OpenStreetMap[9]. The network data from OpenStreetMap includes link length, free flow speed, link type (ex. truck, pedestrian, or motorway), and number of lanes. However, the network data does not include signal timing, which MatSIM does not consider during the simulation. Thus, the travel time or traffic flow from MatSIM might be smaller than the real world if the system is not calibrated. One way to calibrate the systems is to change the free flow speed, as was done in Röder et al. (2013).

This study will aim to use a subnetwork to shorten the simulation time. The subnetwork will be modified to release the congestion from the work zone, such as by changing the signal timing. Although VISTA and MatSIM both can optimize vehicle routing,

[8] Moekel (2016), https://www.mobil-tum.vt.bgu.tum.de/fileadmin/w00bqi/www/mobilTUM2016/Conference_Proceedings/Session_2/0_Moeckel.pdf.
[9] OpenStreetMap website: https://www.openstreetmap.org/#map=5/51.500/-0.100

MatSIM cannot simulate signal timing. Thus, this study will use VISTA as the DTA simulator tool.

2.4 Microsimulators: CORSIM and VISSIM

A time-based microscopic traffic simulator models every vehicle movement every time increment based on car-following and inter-vehicle models, such as lane changing and gap acceptance. Thus, this type of simulator can provide detailed reference data. Two of the most popular microscopic traffic simulators are CORSIM and VISSIM.

The Federal Highway Administration (FHWA) developed CORSIM, a traditional link-node based network model in which links and nodes represent roads and intersections that include signal controllers. This type of network is very similar to reality. In addition, CORSIM is a link-based model that shows the average link travel time, but because CORSIM provides extension programing, it can track individual vehicle position every second and path travel times.

VISSIM was developed by the University of Karlsruhe, Germany during the 1970s and commercialized at PTV Planung Transport Verkehr AG in Germany in 1993 (Bloomeberg et al., 2000). VISSIM is a link-connector based network model, in which links represent roads and connectors represent link junctions. The concept of link junctions is approximate to the concept of nodes. Bloomeberg et al. (2000) indicated that connectors have more flexibility in simulating and controlling different traffic conditions, such as yield signs. Furthermore, VISSIM is a path-based model meaning it shows the average travel time for paths.

Both of these simulators are similar in terms of the inputs required and the output displays. The inputs required for both micro-simulators are a network, traffic demand, and signal control. Traffic demand can be specified as the entry link/nodes traffic volumes and the intersection turn movements, or as an OD matrix. If the user chooses to specify traffic volume using an OD matrix, then the traffic volume will be based on static traffic assignment rather than dynamic traffic assignment. Also, the output of the results can be shown as an animation in both CORSIM and VISSIM. In fact, VISSIM has the capability to provide three-dimensional animation.

Additionally, these simulators also provide a connection to optimization tools for signal control. CORSIM connects to Synchro for the optimization of signal timing, and VISSIM has a component, signal state generator (SSG), which analyzes the impact of traffic conditions. However, Synchro and SSG do not work with actuated, adaptive, transit priority, or ramp metering (Bloomeberg et al., 2000).

Although both simulators have similar structures and capabilities, the ways in which they achieve their functions are different. This study will choose an appropriate microscopic simulator by analyzing its features, including *models, performance,* and *efficiency,* to meet the objective of automatically building a micro-simulation network from the network in VISTA.

Models

CORSIM has two predecessor models, NETSIM and FRESIM; NETSIM models urban streets and FRESIM models freeways. Since the models are not completely integrated yet, the OD traffic specification using the built-in traffic assignment is only available with urban streets through NETSIM.

Conversely, VISSIM does not separate the urban and freeway system. VISSIM also provides three link types: freeway, softcurve, and hardcurve. These links types merely affect the driving behaviors (Horowitz et al., 2005). But unlike CORSIM, the link types do not affect traffic assignment in VISSIM, so VISSIM provides more complete traffic assignment application.

Performance

CORSIM has been developed and calibrated for 40 years in the United States. Default parameters tend to produce typical U.S. traffic conditions even without calibration. On the other hand, VISSIM has more flexibility to modify parameter settings, so the user can tune the model to match reality if effort is expended gathering field data. For example, to simulate I-210 in a congested situation in VISSIM, Gomes et al. (2004) modified parameters describing: necessary lane change, look-back distance, emergency distance, waiting time before diffusion, vehicle following behavior, bumper-to-bumper distance (CC0 and CC1) and the influence between leader and follower (CC4 and CC5).

The primary difference between the software packages is car-following and gap acceptance logic (Bloomeberg et al., 2000). The car-following model (Pitt Model) in CORSIM characterizes a desired amount of headway for individual drivers, but in VISSIM this is replaced by a psycho-physical driver behavior model developed by Wiedemann in 1974. To compare the performance between these simulators, Bloomeberg et al. (2000) analyzed throughput and intersection LOS in a road section of Seattle SR 509 during congestion. He reported that the performance as characterized by throughput and LOS predictions are similar and better than predictions from the HCM because HCM cannot consider the impact of queue spill-backs.

However, Choa et al. (2004) found that CORSIM performance at ramp junctions on freeways was inaccurate in high traffic demands because CORSIM only provides the link density rather than lane density. Lane density is necessary to determine which freeway lanes are directly affected by the ramp. Looking at link density only may minimize the actual impact of traffic conditions associated with ramps. In addition, Choa et al. (2004) analyzed LOS for on-ramp, off-ramp, and weaving sections in the peak hour and found on-ramps to be similar across simulators. However, off-ramps and weaving sections showed different performance as CORSIM respectively showed LOS B and C, and VISSIM showed D for both. Lane density from CORSIM can be obtained through extension programming. The CORSIM Data Dictionary (ITT Industries, Inc. et al., 2006a) describes the parameters and the CORSIM User Guide (ITT Industries, Inc. et al., 2006b) provides instructions to install an extension program for CORSIM. In the CORSIM Data Dictionary, the parameter, VLANE (ID), shows the Lane ID for a specific vehicle ID input. The calculation for lane density is to divide the number of the vehicles occupying the lane by the link length. Although using an extension program could provide detailed traffic data, the process is time consuming.

Efficiency

Since VISSIM provides more opportunities to modify parameters, the time effort for building a network in VISSIM might be more than for building a network in CORSIM. Also, VISSIM is not as user-friendly as CORSIM. Choa et al. (2004) indicated that CORSIM has shorter set up time and VISSIM requires an additional day to refine the network. Even if the user were to use default parameters in VISSIM, CORSIM would be easier to set-up and use.

Coding

CORSIM provides a user-friendly graphical input processor, however, the CORSIM Reference Manual (ITT Industries, Inc. et al., 2001) describes how to create and edit the input data files, so the graphical user interface is not needed for building the network. On the other hand, VISSIM does not provide the same level of guidance for creating and editing the input files, so coding a network might be easier in CORSIM.

Signal Control

CORSIM provides pre-timed and actuated controllers. For non-traditional signal controllers, CORSIM provides a Run Time Extension (RTE) option and instructions in the reference manual on how to code a real time control scheme such as bus priority signals based on the traffic conditions. CORSIM also includes an add-in tool for users to develop NETSIM models. This tool is programmed in C++, and all externally controlled signals should be developed in this same program. Instructions for setting up the NETSIM model and necessary parameters are described in detail in the CORSIM Data Dictionary (ITT Industries, Inc., 2006a) and TSIS User's Guide (ITT Industries, Inc., 2006c).

On the other hand, VISSIM can work with "adaptive" or real-time signal optimization systems. Some of the options available in VISSIM include: SCATS, SCOOT, and VAP (Vehicle Actuated Phasing), etc., for adaptive traffic control. SCATS and SCOOTS both are programs that can adjust green time and cycle according to real time traffic conditions[10]. VAP allows users to install their own signal logic with C or C++ programming according to loop-detector measurement access.

CORSIM and VISSIM Comparison

This study will develop an automated subnetwork creation process that will extract a subnetwork from a larger VISTA network and convert the VISTA subnetwork into a CORSIM-compatible subnetwork. VISTA will provide each node location as needed to build the network in CORSIM.

[10] SCATS website: http://www.scats.com.au/, SCOOT website: http://www.scoot-utc.com/

The other possible micro-simulator choice, VISSIM, uses link-connectors to build the network. This characteristic requires additional steps to create a compatible network from VISTA. Also, CORSIM default parameter values can usually produce approximate real US traffic conditions. Although VISSIM provides users with many potentially useful calibration parameters, users must acquire enough field data and have enough available time and effort to calibrate properly. If users do not have time or data to do the calibration, the calibration ability in VISSIM cannot be used. Therefore, automatically building a network in CORSIM becomes more achievable than in VISSIM.

2.5 CPU Performance

Running a DTA network is time consuming, but the total computational time will ultimately depend on the computer's ability. There are many factors that affect computational performance. The most influential components to computational performance is the computer's central process unit (CPU) and random-access memory (RAM). The RAM is data storage for reading and writing data at almost the same time. The RAM size and the data reading or writing speed affect the computational performance. The CPU focuses on the computation, and the calculation speed depends on the number of cores, the bit size (32 bits or 64 bits), the clock speed (Unit: MHz or GHz), and the internal architecture.

Ideally, the computational time for dual cores is half of what it would be for a single core. But, a 64-bit processor can support 4G+ of RAM while a 32-bit processer only support less than 4G RAM. Therefore, if the RAM is more than 4G, using 64-bit processor performs faster. But if the RAM is less than 4 G, there is no difference between using 32-bit and 64-bit processors.

The clock speed and the internal architecture in a CPU are the biggest factors of the computational performance. The faster the clock, the more instructions the CPU can execute per second. The first one was Intel 8080 CPU which performed 2 MHz in 1974[11], Intel released 500 MHz in 1999 and 1.3 GHz in 2001 (250 times and 650 times

[11] Computer Hope (2017), Computer processor history, available at https://www.computerhope.com/history/processor.htm

than 2 MHz), AMD released 2 GHz in 2004, and until 2013 the highest clock rate was 5.5GHz (2750 times than 2MHz) from IBM.

The clock speed has grown slowly in the past 10 years. The main reason is the advancements made in transistors, which is the main part of a CPU. The transistors switch on and off in order to construct logical gates, and the switch frequency is limited by performing without getting failure[12]. Also, the faster frequency results in more heat, and the CPU might be destroyed if there is not a proper cooling mechanism. Another way to improve computational performance is through the internal architecture, a newer architecture requires less cycles in the same clock speed. Thus, the computational time is less even when the clock speed is identical.

This study cannot find a general way to define the computational time without considering the computer ability because there are too many factors that affect the computational time. Therefore, this study will use seconds or minutes as the unit measuring computational cost, and provide the specifications of CPU used for reference.

2.6 Summary

There are three main ways to allocate traffic, all-or-nothing, system optimal (SO), and user equilibrium (UE). UE is the most realistic of the three types of traffic assignment because it assumes all drivers with the same ODs and departure times seek the shortest travel time path.

Within these types of traffic assignment there are two ways of treating traffic conditions. One way is to assume traffic conditions stay the same throughout the duration of the simulation and the other takes into account that traffic changes over time. DTA is more realistic than STA because STA has three constraints: the traffic conditions are static; STA does not generally restrict link flow to be less than the capacity; STA equalizes the inflow and outflow. Thus, STA output is similar to demand rather than traffic flow. Conversely, DTA traffic conditions are dynamic. The flow is

[12]Mattsson(2014), Why Haven't CPU Clock Speeds Increased in the Last Few Years?, https://www.comsol.com/blogs/havent-cpu-clock-speeds-increased-last-years/

affected by the upstream outflow and downstream inflow; thus, the inflow and outflow are not equal in DTA.

Although DTA provides realistic results, the computational time is much longer than when using STA. Thus, the idea of using subnetworks was proposed as a potential solution. Chen et al. (2012) defined the subnetwork size by size order, i.e. the number of downstream orders. Gemar (2013) developed a process for extracting a subnetwork using ArcGIS, SSH, and VISTA. Also, Bringardener (2015) developed RMSE models that estimated the optimal subnetwork size according to the user defined acceptable error.

This study analyzed two DTA simulators, VISTA and MatSIM and decided to use VISTA as the simulation tool because this program provides a signal control option. Also, this study investigated two traffic simulators, CORSIM and VISSIM, considering performance, and efficiency. In addition, the VISTA network can be easily transformed into a CORSIM network because both simulators are link-based applications.

Lastly, the computational time will undoubtedly depend on the computer ability. There are many factors that affect a computer's computational power. Finding a general way to define the computational time (for example, only providing the clock speed) which can be easily applied to all computers is outside the scope of this study.

CHAPTER 3. DATA

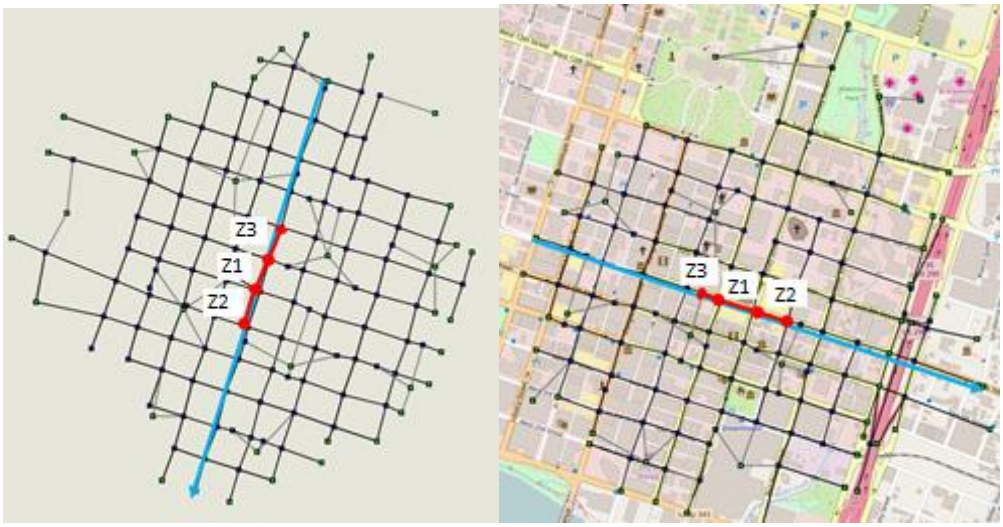
This chapter describes the three networks and the network configurations used in the various data gathering simulation runs. This study chose three locations, Guadalupe Street, 7th Street, and West Martin Luther King Jr. Blvd. (MLK), for building the subnetworks in the Downtown Austin network. The number of links in the sub-area were changed with the subarea's order (7, 9, and 11 order) as well as the number of impacted links (one, two, and three impacted links). In addition, each network has four scenarios (base, 25%, 50%, and 75% capacity reduction), and each scenario has ten simulations. The number of links in the arterials from their location are in Table 3. Overall, there are 5,120 data points in the base scenarios and 15,360 in the impacted scenarios.

Table 3 shows the number of the links for the sub-area size and work zone sizes in Guadalupe, 7th, and West MLK.

Sub-area size order	**7**			**9**			**11**		
Number of impacted(work zone) links	***1***	***2***	***3***	***1***	***2***	***3***	***1***	***2***	***3***
Guadalupe	14	19	21	15	19	21	16	20	22
7th	15	17	19	15	17	19	16	18	19
West MLK	17	20	24	18	21	24	19	23	24

Figure 10 (a) through (c) describe work zone placement with increasing work zone size on Guadalupe Street, 7th Street, West MLK. The Guadalupe Street and 7th Street networks are completely made up of a grid network, whereas West MLK's network has only half of its network as a grid network. Thus, the work zone's traffic impact on the target links in the West MLK network might not be similar to the Guadalupe Street and 7th Street networks. Adding in the results from the West MLK network into the mix will make the regression model more powerful and robust by including a wider variety of outcomes.

Table 4 shows the average traffic flow and v/c across the number of work zones and the sub-areas of different size orders in the base scenarios. Usually, the links on the same arterial have the similar traffic flow and the consecutive work zones' traffic flows are also similar. When a link's capacity is reduced, it is reasonable to have similar traffic reduction on the arterial. The work zones' traffic flow is similar on Guadalupe and 7th Street. Thus, the target link's traffic reduction due to the work zone might be similar. However, the traffic flow and v/c ratios on West MLK are not similar to those found on Guadalupe and 7th Street. The traffic flow in Z2 can be anywhere between 1/3 to 2/3 less than in Z1, depending on the network size order (The bold in Table 4). Figure 10 (c) shows that only 1/3 vehicles stay on West MLK Street after passing the first work zone, and the other 2/3 of the vehicles turn on to Guadalupe Street. Therefore, the traffic reduction on the target links might not be similar when the traffic flows on the work zone are not similar to those on the arterial.



(a) (b)

(c)

Figure 10 The locations of the work zones and the arterials (a) Guadalupe Subnetwork (b) 7th Street Subnetwork (c) West MLK Subnetwork

Table 4 The individual and average work zone traffic flow across the sub-area size, work zone size, and networks

Sub-area Size	Work Zone Size (links)	**Work Zone**	Guadalupe Network		7th Street Network		West MLK Street Network	
			flow	v/c	flow	v/c	flow	v/c
7	one	Z1/Average	955	0.16	1219	0.20	1094	0.36
	two	Z1	947	0.16	1216	0.20	1098	0.37
		Z2	790	0.13	1196	0.20	**359**	0.12
		Average	869	0.14	1206	0.20	729	0.24
	three	Z1	990	0.17	1261	0.21	1060	0.35
		Z2	836	0.14	945	0.16	**343**	0.11
		Z3	775	0.13	1216	0.20	781	0.26
		Average	867	0.14	1141	0.19	728	0.24
9	One	Z1/Average	939	0.16	1184	0.20	1038	0.35
	Two	Z1	946	0.16	1166	0.19	1075	0.36
		Z2	786	0.13	1173	0.20	**390**	0.13
		Average	866	0.14	1170	0.19	733	0.24
	Three	Z1	981	0.16	1229	0.20	1063	0.35
		Z2	805	0.13	875	0.15	**679**	0.23
		Z3	742	0.12	1192	0.20	801	0.27
		Average	843	0.14	1098	0.18	848	0.28
11	One	Z1/Average	959	0.16	1186	0.20	1057	0.35
	Two	Z1	962	0.16	1166	0.19	1048	0.35
		Z2	806	0.13	1173	0.20	**328**	0.11
		Average	884	0.15	1170	0.19	688	0.23
	Three	Z1	946	0.16	1219	0.20	1059	0.35
		Z2	793	0.13	880	0.15	**684**	0.23
		Z3	710	0.12	1191	0.20	798	0.27
		Average	752	0.13	1036	0.17	741	0.25

Summary

This study chose three locations, Guadalupe Street, 7th Street, and West Martin Luther King Jr. Blvd. (MLK), for building the subnetworks in the Downtown Austin network. The number of the links changed by the sub-area size order and the number of impacted link at each location. In addition, each subarea simulated four scenarios, the base, 25% capacity reduction, 50% capacity reduction, and 75% capacity reduction. Overall, there are 5,120 data points in the base scenarios and 15,360 in the impacted scenarios.

The network configurations at leach location are: Guadalupe Street is a grid network, the work zone's v/c in the base scenario is lowest, and the v/cs are similar across the three work zones; 7th Street is a grid network and the work zone's v/c is higher than the Guadalupe Street, and the v/c are also similar across the three work zones; West MLK's network has only half of its network as a grid network, the work zone's v/c is highest, but the Z2's v/c is only between 1/3 and 2/3 less than in Z1.

CHAPTER 4. OPTIMAL SUB-AREA SIZE AND SIMULATION TIME

Although using dynamic traffic assignment (DTA) can provide a more realistic result compared to static traffic assignment, it is very time consuming for DTA to reach user equilibrium. This long simulation time can seem very inefficient when considering that some network changes merely impact a portion of the network. Thus, researchers (Bringardener et al., 2015; Bringardener, 2014; Gemar et al., 2014; Gemar, 2013; Chen, 2012) have introduced methodology aiming to reduce simulation time through the use of a subnetwork rather than a regional network. However, few studies have focused on quantifying simulation time savings as a result of using a smaller network.

In addition, while using a subarea to represent an entire network is a way to reduce computational time, how to effectively balance simulation error and simulation time remains a key question. When using a subnetwork, the hope is to capture all the significant impacts resulting from a change in the network. For example, suppose a single lane on a single block is closed-off within the central area of a network. It is very likely that this single blockage will not affect travelers that are very far away from the blockage and that the blockage will only affect a sub-area. In fact, that's good news because running a full network in a DTA simulation can last an unbearably long amount of time. By extracting a sub-area of a network, the simulation time can be cut down. If the subnetwork size is very small, the simulation time can potentially be reduced by a lot, but the result might not be anywhere near the result one would get from running the entire network (i.e., high error, short simulation time). On the other hand, if the subnetwork size is very large, the simulation time may not be significantly reduced (i.e. very low error, very long simulation time).

Thus, this study has investigated the key indexes affecting simulation time to help quantify computational effort required. Also, this study has developed models for defining an optimal DTA subnetwork size. Two objective functions have been developed to define the balance between the error and the simulation time, and the constraints followed Bringardener (2015) RMSE models to define the sub-area size and control error.

4.1 Simulation Time Estimation

The pattern of a network can potentially influence simulation time. There are different types of network patterns, such as grid or radial (Figure 11). Levin et al. (2015) compared the number of the simulation iterations to converge by simulating Anaheim, with 52,347 trips, Downtown Austin, with 89,078 trips, and the Williamson County network, with more than 201,588 trips. Downtown Austin is a grid network, and the other two networks are more spread out. Although Downtown Austin performed worst in the first ten iterations, this network converged faster than the other two networks. These authors thought the reason that Downtown Austin performed best was this network was a grid-like network, which is simpler and more uniform in pattern, leading to the conclusion that the network pattern can influence simulation time.

One of the factors influencing the simulation time is the computer speed. The CPU used for simulating VISTA is 3.33 GHz and 24 cores. This study cannot find a general algorithm that can transform the computational time to units which are independent computer characteristics, so the unit of the computational cost used is still seconds.

Figure 11 Example of the grid[13] and the radburn[14] network pattern (The left is the grid, and the right the is radburn pattern)

This study only focused on the grid-type Downtown Austin network to investigate the key indexes. Three networks were chosen, Guadalupe Street, a grid-type network, 7th Street, a grid-type network and close to the freeway, and West MLK, half grid-type

[13] Wu (2007), *How PAT Will Shape The City In The Future* ,http://www.acroscape.com/PAT.html
[14]Phillips Preiss Grygiel LLC (2015*), Development and Preservation Strategies - Radburn, Fair Lawn, NJ*, http://www.ppgplanners.com/development-and-preservation-strategies-radburn-fair-lawn-nj.html

network. Each network was simulated by changing capacity reduction, sub-area size, and work zone size. There are 1080 data points that where indexed in the following way,

- demand: The number of the vehicles in the sub-area;
- links: The number of the links in the sub-area;
- size: The connected order from the work zone to the most faraway link in the sub-area; (See Chapter 5)
- zone: The number of impacted links in the sub-area;
- reduction: The work zone capacity reduction;

Figure 12 shows a linear relationship between demand and simulation time. This result is reasonable because as the number of vehicles increase, the simulation will require more computational time to calculate the paths for the additional vehicles.

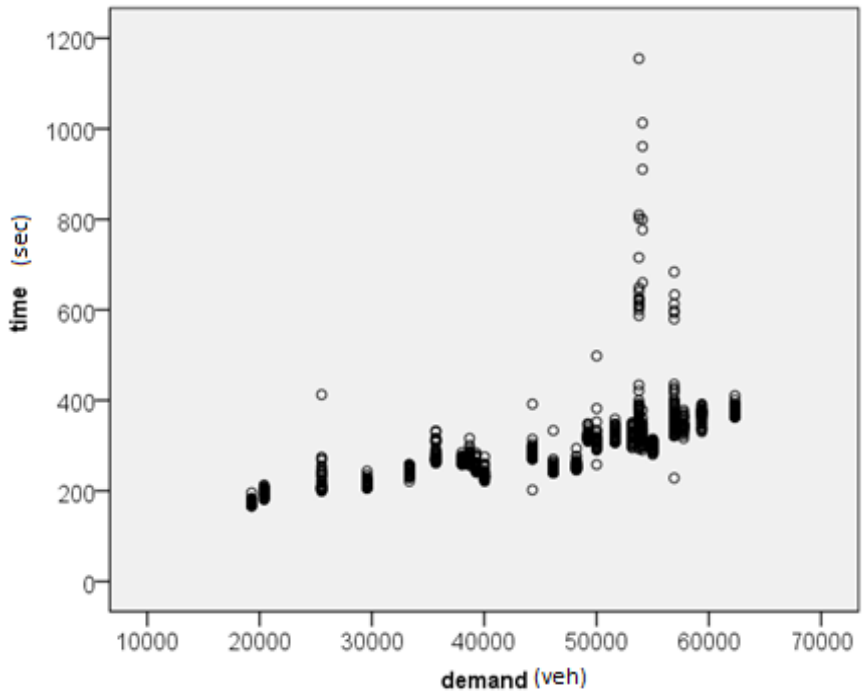


Figure 12 Relationship between demand and simulation time

Figure 13 shows that there is a linear relationship between the number of links and simulation time. VISTA analyzes vehicle travel times using the cell transmission model, so more links means calculating more cells. As the number of cells increases, the computational time also increases.

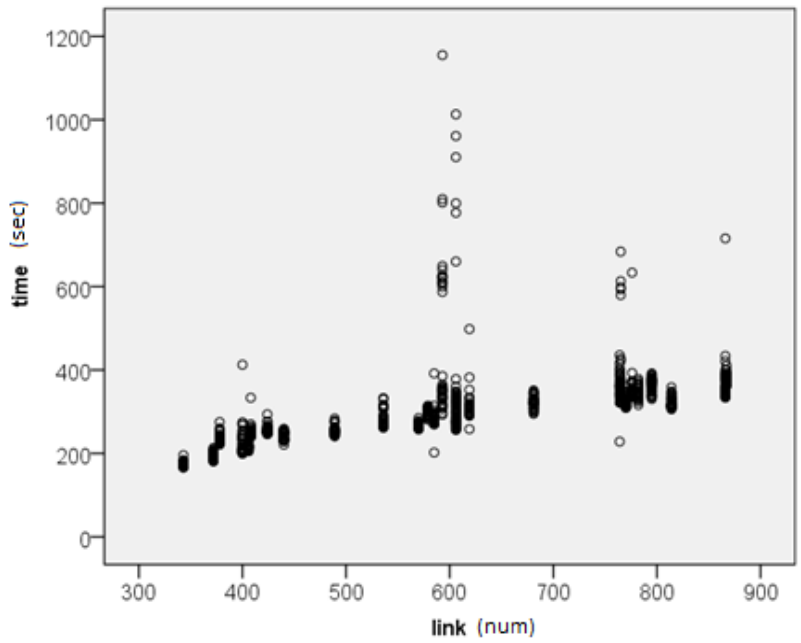


Figure 13 Relationship between number of links link and simulation time

Figure 14 shows a linear relationship between size order and simulation time. However, the simulation time is more diverse when the network size is 9. Thus, defining the simulation time might not be so easy when the sub-area size is 9.

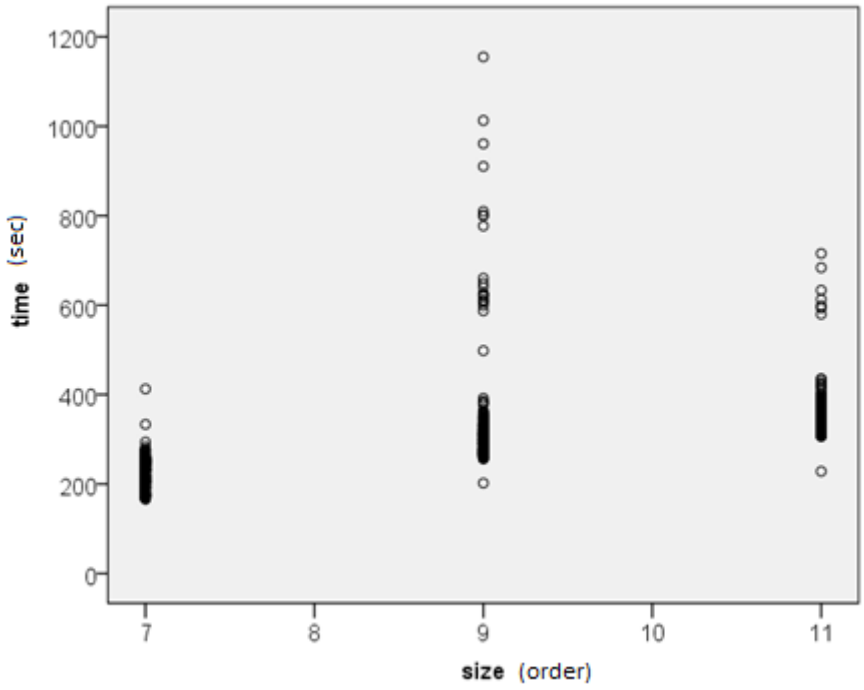


Figure 14 Relationship between size order and simulation time

Figure 15 shows the relationship between zone (number of impacted links) and simulation time. However, it is hard to define the relationship between these variables from this figure. Thus, this study will use Pearson Correlation to analyze the relationship (Table 5).

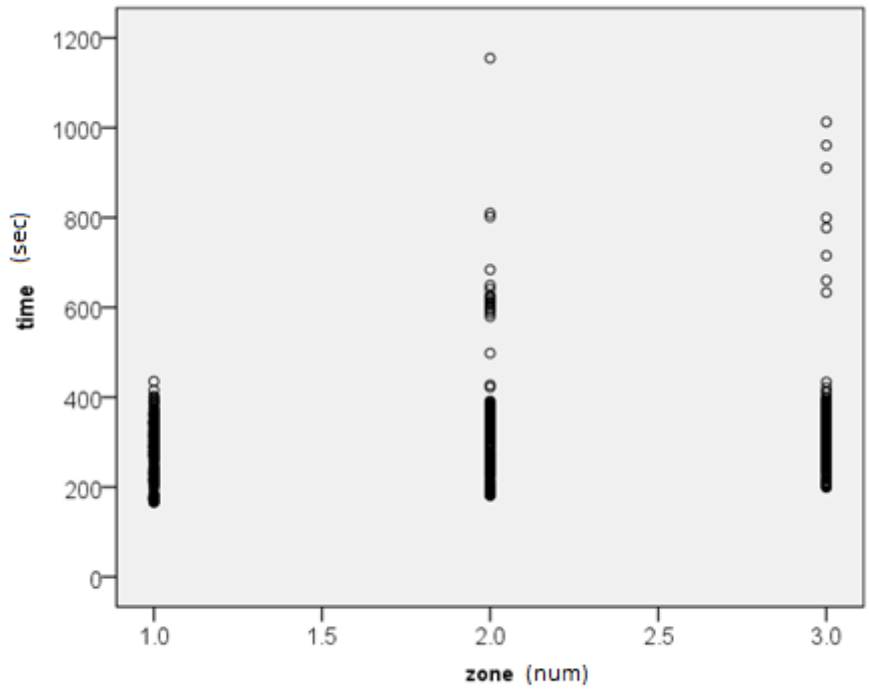


Figure 15 Relationship between zone (number of impacted links) and simulation time

Figure 16 shows the relationship between the work zone capacity reduction and simulation time. However, it is hard to define the relationship between these variables from this figure as well.

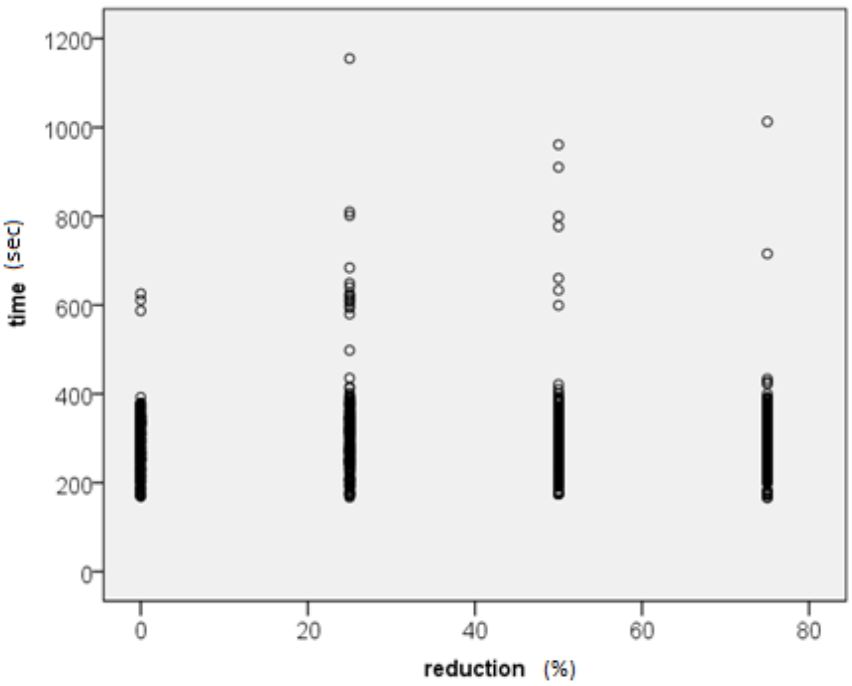


Figure 16 Relationship between zone capacity reduction and simulation time

According to Figure 12 through Figure 16, we can see there is a linear relationship between simulation time and demand, link, and size indexes. However, it is hard to verify the relationship between the simulation time and the zone and reduction. Linear regression assumes all the control variables are independent. Therefore, these indexes should not correlate otherwise only one index should be used. Table 5 shows the correlation between these indexes and the simulation time using Pearson Correlation coefficient.

Table 5 The correlations between the indexes and simulation time

		zone	size	link	demand	reduction	time
zone	Pearson Correlation	1	.000	.157**	.168**	.000	.164**
	Sig. (2-tailed)		1.000	.000	.000	1.000	.000
	N	1080	1080	1080	1080	1080	1080
size	Pearson Correlation	.000	1	.972**	.772**	.000	.580**
	Sig. (2-tailed)	1.000		.000	.000	1.000	.000
	N	1080	1080	1080	1080	1080	1080
link	Pearson Correlation	.157**	.972**	1	.817**	.000	.596**
	Sig. (2-tailed)	.000	.000		.000	1.000	.000
	N	1080	1080	1080	1080	1080	1080
demand	Pearson Correlation	.168**	.772**	.817**	1	.000	.650**
	Sig. (2-tailed)	.000	.000	.000		1.000	.000
	N	1080	1080	1080	1080	1080	1080
reduction	Pearson Correlation	.000	.000	.000	.000	1	.022
	Sig. (2-tailed)	1.000	1.000	1.000	1.000		.476
	N	1080	1080	1080	1080	1080	1080
time	Pearson Correlation	.164**	.580**	.596**	.650**	.022	1
	Sig. (2-tailed)	.000	.000	.000	.000	.476	
	N	1080	1080	1080	1080	1080	1080

**. Correlation is significant at the 0.01 level (2-tailed).

From Table 5, except for capacity reduction, all the variables are correlated to simulation time. Also, demand (Adjusted R Square= 0.65), has the best fit to the simulation time. However, demand is correlated with the other indexes, except for reduction. In order to develop a linear regression model for the simulation time using demand, this model cannot use the other related indexes. The linear regression model is shown in Table 6.

Table 6 Statistical result of predicting simulation time using demand

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
0.650	0.422	0.421	68.132

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	3653208.431	1	3653208.431	786.986	.000
Residual	5004103.505	1078	4642.026		
Total	8657311.936	1079			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	80.310	8.180		9.817	0.000
demand	0.005	0.000	0.650	28.053	0.000

Zone and size are both correlated with simulation time, but are not correlated with each other. Unlike link and demand, both indices are defined by the practitioner. Therefore, there is no need to take the extra step of extracting the sub-area to get these indices. The second regressions model developed uses zone and size and the results are shown in Table 7.

Table 7 Statistical result of predicting simulation time using zone and size

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.656[a]	0.430	0.428	67.716

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	3723346.569	3	1241115.523	270.663	.000[b]
Residual	4933965.367	1076	4585.470		
Total	8657311.936	1079			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-161.452	17.844		-9.048	0.000
size	32.467	1.263	0.592	25.704	0.000
zone	30.244	2.751	0.276	10.993	0.000
base flow	0.118	0.011	0.281	11.187	0.000

The adjusted R square is higher in Table 6 than in Table 7. Therefore, this study recommends using demand as the index for predicting the simulation time if the practitioner can easily extract the sub-area. However, if extracting a sub-area is time consuming, around 20 minutes following Bringardner's process and around 5 minutes following the auto program(see Bringardner, 2015 for the way to extract a sub-area manually, and see Chapter 6 for the way to extract a sub-area with the auto program), practitioners can predict the simulation time by using size and zone.

Summary

This study investigated the indices influencing simulation time most. These indices include the number of vehicles, the number of links, the network size order, work zone size, and the capacity reduction.

The relationship between simulation time and number of vehicles, the number of links, and the sub-area size was found to be linear. However clear relationships between the rest of the indices and simulation time were not identified.

This study used the Pearson Correlation coefficient to verify the relationship between the indices and simulation time. Only capacity reduction does not significantly correlate with simulation time. Since using the number of vehicles (demand) provides the best result, this index was used to develop a linear regression model for predicting simulation time. The adjusted R square is 0.422. However, to obtain the demand the user must extract the sub-area (See Chapter 6) and define the sub-area size and zone size. Since zone size and the sub-area size are not correlated, this study also developed a linear regression model for the simulation time using these indices. The adjusted R square is 0.428.

Comparing both adjusted R squares leads to the conclusion that predicting simulation time using the number of vehicles (demand) works best. However, this index is more difficult to get without the user defining the sub-area size and the work zone size, and requires the extra step of extracting the sub-area. Thus, this study would recommend practitioners use the number of vehicles as the index if it is not too time consuming to extract the sub-area. Otherwise, practitioners can predict the simulation time using the zone and size indexes.

4.2. Objective Functions and Constraints

Gemar (2013) provided a table showing the largest vehicle rerouting area according to work zone size and capacity reduction. Also, Bringardner (2015) developed two models to estimate the largest rerouting area, and used an acceptable error approach to estimate the acceptable sub-area size. In this study, the network size that provides the largest rerouting area is called the perfect size, meaning that the size can capture all the vehicle rerouting from the work zone in question. The perfect sub-area can represent the entire network's performance because it captures all the significant impacts. The estimated simulation time for the perfect size is called perfect time.

However, the perfect size network can still be very large and the perfect time may still be too long for some study timelines; thus, the optimal sub-area size was created to define the balance between the computational time and the error. Also, the optimal sub-area size's computational time is called optimal simulation time. The optimal sub-area size and time will be smaller than the perfect sub-area size and time.

Two objective functions were developed,

$$\text{Min } Z1 = \left(1 - \frac{Size_{optimal}}{Size_{Perfect}}\right) + \alpha * \frac{Time_{Optimal}}{Time_{Perfect}}, \qquad (13)$$

$$\text{Min } Z2 = \frac{Size_{Perfect}}{Size_{Optimal}+e} + \alpha * \frac{Time_{Optimal}}{Time_{Perfect}}, \qquad (14)$$

where $Z1$ and $Z2$ are the objective functions, $Size_{optimal}$ is the optimal subnetwork size, the size is at the balance between the error and the simulation time, $Size_{perfect}$ is the perfect subnetwork size, the size does not have the error, $Time_{optimal}$ is the simulation time at the optimal subnetwork size, $Time_{perfect}$ is the simulation time at the perfect subnetwork size, α the weighting factor for the simulation time, and e is a non-zero positive value.

In both of the objective functions, the first part represents the ratio between the subnetwork sizes. This ratio also implies the error. The error decreases as the optimal subnetwork size approaches the perfect subnetwork size. The second part represents the ratio between the simulation times. When the optimal subnetwork size is close to the perfect subnetwork size, the first part of the equation decreases and the second part

increases. Otherwise, the first part increases and the second part decreases. This study aimed to find the minimum result for the both objective functions.

For the first objective function, Z1 equals α when the optimal subnetwork size equals the perfect size. Otherwise, Z1 is close to 1 when the optimal subnetwork size is 0. For the second objective function, Z2 is $\alpha + 1$ when the optimal and the perfect sizes are equal. Otherwise, Z2 is infinity when the optimal size is 0. This study will compare the objective functions in the next section.

Also, the constraints are as following,

$$Impact_{Perfect_{RMSE}} = 0.430 * Size_{perfect} + 0.135 * R + 0.719 * N - 0.117 * \frac{C}{1000} - 9.610 * v\_over\_c - 0.016 * Size_{perfect} * R + 0.145, \quad (16)$$

$$Base_{Perfect_{RMSE}} = -0.115 * Size_{perfect} - 0.018 * C - 15.487 * v\_over\_c + 6.004, \quad (17)$$

$$Perfect_{Time} = -161.452 + 32.467 * size_{perfect} + 30.244 * N + 0.118 * BaseFlow, \quad (18)$$

$$Optimal_{Time} = -161.452 + 32.467 * size_{optimal} + 30.244 * N + 0.118 * BaseFlow, \quad (19)$$

$$Base_{Perfect_{RMSE}} = Impact_{Perfect_{RMSE}}, \quad (20)$$

$$Size_{optimal} \leq Size_{perfect}, \quad (21)$$

where R is percent capacity reduced, N is the number of impacted links/work zone size, C is work zone average capacity v_over_c the volume to capacity ratio, and $BaseFlow$ is work zone average flow in base scenarios.

Equation (16) through (18) define the perfect subnetwork size and time, and the optimal subnetwork size time. This study referred to Bringardener (2015) to define the perfect subnetwork size. This author developed the base and impacted regression models ($Base\_Perfect\_RMSE$ and $Impact\_Perfect\_RMSE$), and defined the perfect

subnetwork size as Equation (20). In addition, this study used the linear regression model from section 4.1 to define the sub-area size simulation time. Thus, the perfect subnetwork size and the simulation time are known variables.

There are two steps for defining the optimal subnetwork size,

1. Determine whether the scenario requires the subnetwork or not.

- No need for a subnetwork size when $Base\_Perfect\_RMSE$ value is less than $Impact\_Perfect\_RMSE$ when network size is 0.
- Cannot find the perfect subnetwork size ($Impact\_Perfect\_RMSE$ always> $Base\_Perfect\_RMSE$).

2. If yes, run the optimal model.

- Define α, the weighting factor for the simulation time.
- Run the models

For the first step, there are two situations in which the base and the impacted models can never equal each other. One is when the base RMSE is larger than the impact RMSE when the sub-area size is 0, which means there is no traffic impact on the network; the other is the impact RMSE is always larger than the base RMSE, which means the traffic impact is so large that we can only use the entire network to capture the significant impact rather than a subnetwork.

4.3 Case Study

A case study was used to compare the two objective functions. For the first objective function, we assumed the C as 6000, and changed v_over_c from 0.2 to 0.3 in increments of 0.05, zone size from 1 to 3, capacity reduction from 25 to 75, and α from 0.5 to 1.5. The result is shown in Table 8. The cells in the table shows the optimal and perfect subnetwork size. The value at the left side of the slash is the optimal size, and at the right is the perfect size. If the cell shows 0, then the $Base\_Perfect\_RMSE$ value is less than the $Impact\_Perfect\_RMSE$ when network size is 0. Therefore, the user does not need to study the traffic impact. If the cell shows entire, $Impact\_Perfect\_RMSE$ is always larger than $Base\_Perfect\_RMSE$. The user needs to use the entire network to study the entire significant traffic impact.

Table 8 Example result for Z1 objective function

	α =1			α =0.5			α =1.5		
v/c=0.2	Work Zone Size			Work Zone Size			Work Zone Size		
Reduction	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
25	0	0	0/8	0	0	**8/8**	0	0	0/8
50	0/9	11/11	14/14	**9/9**	11/11	14/14	0/9	**0/11.**	**0/14**
75	0/8	10/10	11/11	**8/8**	10/10	11/11	0/8	**0/10**	**0/11**
v/c=0.25	Work Zone Size			Work Zone Size			Work Zone Size		
Reduction	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
25	0	0	entire	0	0	entire	0	0	entire
50	0/10	13/13	15/15	**10/10**	13/13	15/15	0/10	**0/13**	**0/15**
75	0/9	10/10	11/11	**9/9**	10/10	11/11	0/9	**0/10**	**0/11**
v/c=0.3	Work Zone Size			Zone Size			Work Zone Size		
Reduction	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
25	0	0	entire	0	0	entire	0	0	entire
50	0/11	14/14	17/17	**11/11**	14/14	17/17	0/11	**0/14**	**0/17**
75	0/9	10/10	12/12	**9/9**	10/10	12/12	0/9	**0/10**	**0/12**

The bold number shows the optimal size different to the optimal size at α =1**.** Table 8 shows that this objective function is sensitive. The number to the left of the / represents the optimal subnetwork size and the number to right of the / represents the perfect subnetwork size. When α is 1, the optimal subnetwork size is either 0 or equal to the perfect subnetwork size. When α is 0.5, all the optimal subnetwork sizes equal to the perfect subnetwork size. However, when α is 1.5, all the optimal subnetworks drop to 0.

For the second objective function, all the parameters are identical to the previous one, except the parameter α is tested with values of 1 to 5. The result is shown in Table 9. When α = 1, all the optimal sizes are equal to the perfect sizes, but when α = 5, the optimal sizes are around half the perfect subnetwork size. Thus, compared to the first objective function, the second objective function has more flexibility in changing the subnetwork size by manipulating α.

Table 9 Example result for Z2 objective function

	α =1			$\alpha = 5$		
v/c=0.2	Work Zone Size			Work Zone Size		
Reduction	**1**	**2**	**3**	**1**	**2**	**3**
25	0	0	8/8	0	0	0/8
50	9/9	11/11	14/14	**4/9**	**5/11**	**7/14**
75	8/8	10/10	11/11	4/8	**4/10**	**5/11**
v/c=0.25	Work Zone Size			Work Zone Size		
Reduction	1	2	3	**1**	**2**	**3**
25	0	0	entire	0	0	entire
50	10/10	13/13	15/15	**4/10**	**6/13**	**7/15**
75	9/9	10/10	11/11	**4/9**	**5/10**	**5/11**
v/c=0.3	Work Zone Size			Work Zone Size		
Reduction	**1**	**2**	**3**	**1**	**2**	**3**
25	0	0	entire	0	0	entire
50	11/11	14/14	17/17	**5/11**	**6/14**	**8/17**
75	9/9	10/10	12/12	**4/9**	**5/10**	**5/12**

This study would like to calculate the percent of the sub-area errors and the computational timing saving for these objective functions. For the percent of sub-area errors, this study referred the concept from Bringardener (2014), the equation of the sub-area error percentage is $\left(1-\frac{Size_{optimal}}{Size_{Perfect}}\right)*100\%$. Also, for the percent of computational timing saving, this study sued Equation (18) and (19), which show the computational time when the sub-area size is able to represent the entire network and the optimal size. The equation of the percent of computational timing saving is $\left(1-\frac{Time_{optimal}}{Time_{Perfect}}\right)*100\%$.

For the first objective function, since the optimal sub-area size either equals to 0 or the perfect size, the sub-area error is either 100% or 0%, and the percent of computational time saving is 100% or 0%. For the second objective function, when $\alpha = 5$, the sub-area error is around 50% when v/c=0.2, 0.25, and 0.3, and the error is between 50% and 60%.

4.4 Summary

This study investigated the factors that influenced simulation time the most. These factors included the number of vehicles, the number of links, the network size order, work zone size, and the capacity reduction. This study found the number of vehicles had the strongest linear correlation with the simulation time (adjusted R square is 0.422).

However, it is difficult to obtain the number of the vehicles without defining the sub-area size and the work zone size. Thus, this study also developed a linear regression model for using the sub-area size, the work zone size, and zone average flow(adjusted R square is 0.428).

Then, this study developed two objective functions for the optimal sub-area size and simulation time. Both objective functions had two parts, the first part represented the error of the sub area size, and the second part represented the ratio of the simulation time. Also, there was a weighting factor for the simulation time. As the optimal subnetwork size increased, the error of the sub area size decreased and the ratio of the simulation time increased. The minimum results from the objective functions were the optimal sub-area size and the simulation time.

The difference between the two objective functions was the way to represent the error in the first part. The first part in the first objective function was $\left(1-\frac{Size_{optimal}}{Size_{Perfect}}\right)$, and in the second part was $\frac{Size_{Perfect}}{Size_{Optimal}+e}$. When the optimal subnetwork size was close to zero, the first part in the first function was close to one but in the second function was infinity.

Finally, a case study was implemented to compare the two objective functions when the weighting factor changes. The result showed that the first objective function was sensitive, the weighted factor was 1±0.5 and the optimal subnetwork size was either equal to zero or equal to the perfect subnetwork size. Conversely, the second objective function had more flexibility in changing the subnetwork size by manipulating the weighted factor. When the weighted factor increased from one to five, the optimal subnetwork size decreased from equal to the perfect subnetwork size to around half of

the perfect subnetwork size. Also, the sub-area error is around 50% and the percent of the computational time saving is between 50% and 60%.

CHAPTER 5. LINEAR REGRESSIONS FOR CORSIM SUB-ARTERIAL SIZE

There are several indexes to identify an impacted area, such as the sub-network's demand, the vehicle rerouting, or the traffic flow difference between base and impacted scenarios. In this study, flow difference is most appropriate because it is the CORSIM required input. The index is affected by two factors, random seeds and number of impacted links. Due to random seeds, traffic flow changes across replicate simulation runs even in a base scenario with no impacted links. Thus, the one way to identify change in the network is by finding an impacted sub-area from the difference between the base and impacted scenarios in terms of traffic flow.

This study aims to develop regression models using data from simulation runs to identify impacted areas. There are several sections in this chapter, including *Methodology*, *Statistical Analysis*, *Boxplots Analysis*, and *Models*. The *Methodology* section defines the dependent variable, a Root Mean Square Error (RMSE) for flow difference between base and impacted scenarios, in the regression model. This section also describes using the Confidence Interval (CI) for identifying an impacted sub-area. The statistical analysis investigates the hypothesis of a regression model. The *Boxplots Analysis* section analyzes the key factors related to defining flow difference, and determines the type of regression models to build in the Models section. The *Models* section analyzes the factors related to the defined flow difference, and develops the regression models between the defined flow difference and other factors.

5.1 Methodology

This study has created regression models that define an appropriate subnetwork size to capture the impacts from a change (i.e., work zone) in the network. Keeping subnetwork size and the work zone size constant, traffic flow is the only variable that changes across base scenario simulations due to the random seeds in the simulator. On the other hand, traffic flow changes in the impacted scenario not only because of the random seeds but also because of the capacity reduction in the impacted links. The question that follows is whether the flow difference between the base and the impacted scenario is because of the random seeds or the capacity reductions. To answer this question, two definitions are required,

1. The traffic difference between the base and the impacted scenarios,
2. The boundary of traffic difference in the base scenario.

For defining flow difference between base and impacted scenarios, this study refers and modifies the RMSE equation from Bringardner (2015) (22). The variables in Bringardner's model are demands, average base demand for all base scenarios, and a subnetwork demand extracted from an individual simulation. In the beginning, this study followed Bringardner (2015) by defining a sub-area in terms of size order. Bringardner used ten simulations for base and impacted scenarios, so that each test network provided ten impact and base RMSE values for each subnetwork size order tested. This study modifies this equation by replacing demand with traffic flow. The modified equation is,

$$RMSE = \sqrt{(f_{ij} - \bar{f_i})^2} \quad (22),$$

where, f is the traffic flow at link i from an individual simulation j, and $\bar{f_i}$ is the average traffic flow at link i from the base scenario. The cumulative link flows are called the order flow Figure 17 shows an example for the order flow. The lighter links are size parameter one in the left figure, and the first order link flow adds the flows for all of the lighter links (adding all xi, i =1 through 6). The darker links are size parameter 2 in the right figure, and the second order flow adds only the dark links for the second order flow (adding all yi, i = 1 through 12). If a sub-area with size parameter of five and ten simulation runs for the base and impacted scenarios provides 50 base (1^{st} order = ten runs, 2^{nd} order = ten runs,...5^{th} order = ten runs) and 50 impact RMSE values. However, it was found that for the same size order some link flow values change drastically, while others do not. Thus, the order flows could potentially not change much between base and impacted scenarios, even though the impacted links have a large capacity reduction.

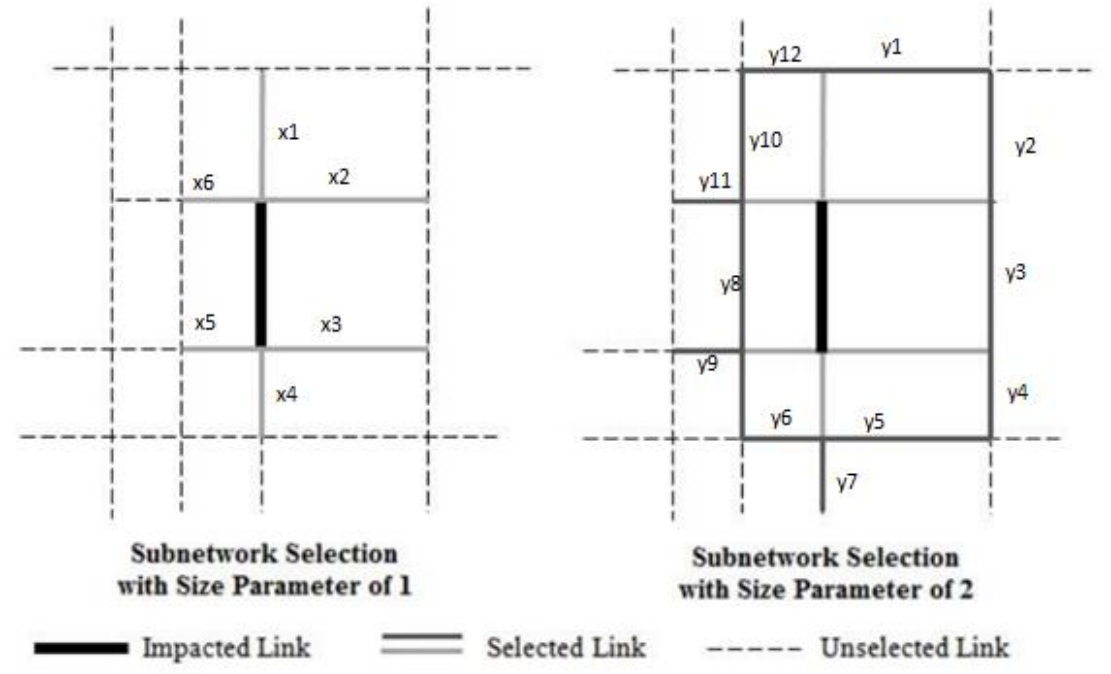


Figure 17 shows an example of the order flow (modified from Gemar, 2013)

For defining the boundary of traffic difference in the base scenario, this study uses the confidence interval,

$$\bar{x} \pm s * t_{\alpha,n-1} * \sqrt{\frac{1}{n}}, \qquad (22)$$

for defining flow difference between base and impacted scenarios where $\bar{x}$ is the mean traffic flow, $t\alpha/2$, $n-1$ is the t-statistic for α level of significance and a sample size of n, and s is the sample standard deviation. This concept was compared to the hypothesis test to determine the robustness of the confidence interval for evaluating whether the impacted scenario traffic flow error measure falls within the expected range of values found in the base scenarios. The confidence interval was used as an approach to evaluate appropriate subnetwork sizes in place of the more time consuming full hypothesis test. The hypothesis tests require simulating the both base and impacted scenario multiple times, but the confidence interval only requires multiple base runs. Since multiple impacted scenarios are likely to be tested for traffic control planning processes, the confidence interval can save a significant amount of time.

5.2 Statistical Analysis

This study uses the confidence interval to define the boundary of base RMSEs. The assumptions of confidence interval are the RMSEs should be random; the RMSEs follow a normal distribution; and the RMSE variances are identical. The first assumption has been accepted due to the random seeds present in the simulation. To verify the rest of the assumptions, this study uses various statistical tests.

Table 10 verifies the data follow a normal distribution and have identical standard deviations. This table summarizes the RMSE statistical test results for the first hour ten base scenario simulations and ten impacted scenario simulations. Since there are 512 links in this network, it would be very time consuming to test every single link, so this study has chosen two links to do the statistical tests in one zone in sub-area size 7, two zones in sub-area size 9, and three zones in sub-area size 11 network as the trial scenarios. Furthermore, the Shapiro-Wilk Test, Levene's Test, and the T-test are used for testing normality, equal variance in the impacted scenario to the base scenario, and identical mean.

Table 10 Statistical Tests of the RMSE for 10 Base and 10 Impacted Scenarios (Simulation time = 3600 sec)

Loc.	Link ID	Sub-area Size	Zone Size	Capacity Reduction	Normality[1]	Equal Variance[2]	Equal Mean[3]
Guadalupe Street	18333	7	1	0	Y	--	--
	18333	7	1	25	Y	Y	Y
	18333	7	1	50	Y	Y	Y
	18333	7	1	75	Y	Y	N
	18470	7	1	0	Y	--	--
	18470	7	1	25	Y	Y	Y
	18470	7	1	50	Y	Y	N
	18470	7	1	75	Y	Y	N
	18333	9	2	0	N	--	--
	18333	9	2	25	Y	Y	Y
	18333	9	2	50	Y	Y	Y
	18333	9	2	75	Y	Y	N
	18470	9	2	0	Y	--	--
	18470	9	2	25	Y	Y	Y
	18470	9	2	50	Y	Y	N
	18470	9	2	75	Y	N	N
	18333	11	3	0	Y	--	--
	18333	11	3	25	Y	Y	Y
	18333	11	3	50	Y	N	N
	18333	11	3	75	Y	N	N
	18470	11	3	0	Y	--	--
	18470	11	3	25	Y	Y	Y
	18470	11	3	50	Y	Y	N
	18470	11	3	75	Y	Y	N

Table 10 Continued. Statistical Tests of the RMSE for 10 Base and 10 Impacted Scenarios (Simulation time = 3600 sec)

Loc.	Link ID	Sub-area Size	Zone Size	Capacity Reduction	Normality[1]	Equal Variance[2]	Equal Mean[3]
7th Street	105209	7	1	0	Y	--	--
	105209	7	1	25	Y	Y	Y
	105209	7	1	50	Y	Y	N
	105209	7	1	75	Y	N	N
	105253	7	1	0	N	--	--
	105253	7	1	25	N	N	Y
	105253	7	1	50	Y	Y	N
	105253	7	1	75	Y	N	N
	105209	9	2	0	Y	--	--
	105209	9	2	25	Y	Y	Y
	105209	9	2	50	Y	Y	N
	105209	9	2	75	Y	Y	N
	105253	9	2	0	Y	--	--
	105253	9	2	25	Y	Y	N
	105253	9	2	50	Y	N	N
	105253	9	2	75	Y	N	N
	105209	11	3	0	Y	--	--
	105209	11	3	25	Y	Y	Y
	105209	11	3	50	Y	Y	N
	105209	11	3	75	Y	Y	N
	105253	11	3	0	N	--	--
	105253	11	3	25	N	N	Y
	105253	11	3	50	N	Y	N
	105253	11	3	75	Y	N	N

Table 10 Continued. Statistical Tests of the RMSE for 10 Base and 10 Impacted Scenarios (Simulation time = 3600 sec)

Loc.	Link ID	Sub-area Size	Zone Size	Capacity Reduction	Normality[1]	Equal Variance[2]	Equal Mean[3]
West MLK	6229	7	1	0	N	--	--
	6229	7	1	25	N	Y	Y
	6229	7	1	50	N	Y	Y
	6229	7	1	75	Y	Y	N
	6227	7	1	0	N	--	--
	6227	7	1	25	N	Y	N
	6227	7	1	50	N	Y	Y
	6227	7	1	75	Y	N	Y
	6229	9	2	0	N	--	--
	6229	9	2	25	Y	N	N
	6229	9	2	50	N	Y	N
	6229	9	2	75	Y	N	N
	6227	9	2	0	Y	--	--
	6227	9	2	25	Y	Y	Y
	6227	9	2	50	Y	Y	Y
	6227	9	2	75	Y	Y	N
	6229	11	3	0	N	--	--
	6229	11	3	25	Y	Y	Y
	6229	11	3	50	Y	N	N
	6229	11	3	75	Y	N	N
	6227	11	3	0	Y	--	--
	6227	11	3	25	Y	N	N
	6227	11	3	50	N	N	N
	6227	11	3	75	Y	N	N

1. Y = Accept H_o: Distribution is normal; N = Reject H_o, conclude H_a: Distribution is not normal
2. Y = Accept H_o: $\sigma_{base} = \sigma_{impact}$; N = Reject H_o, conclude H_a: $\sigma_{base} \neq \sigma_{impact}$
3. Y = Accept H_o: $\mu_{base} = \mu_{impact}$; N = Reject H_o conclude H_a: $\mu_{base} \neq \mu_{impact}$

Table 10 can be interpreted using the first five columns to establish the scenario: location by street name; the target link's id; subnetwork selection size parameter; impacted size in number of links; and impacted magnitude in percent capacity reduction. As seen in Table 10, the majority of scenarios passed the normality and equal variance tests. The last column displays target link's RMSE mean will become different from the base scenario mean when capacity reduction increases.

5.3 Boxplot Analysis

For a particular network configuration, an appropriate sub-network size can be determined by statistical analysis, but it may take major computational time and effort. Specifying this relationship may help characterizing an appropriate sub-area with significantly less time and effort. This study aims to predict the confidence interval and the predicted impact RMSE value for identifying an impacted subarea. We assumed that the target links would experience significant impact from the work zones when the target link's confidence interval is less than the predicted impact RMSE value. Thus, two regression models have been built, one for the base and one the impact, for forecasting the confidence interval and the RMSE in the impacted scenario.

Boxplots were created to describe RMSE values with respect to changing capacity reduction, work zone size, and sub-area size. Since there are three locations and 81 networks, three networks ranging from the extreme to the middle type are described. The following list describes the chosen networks for analyzing each of following variables,

1. *Capacity reduction*: network size 7 and one work zone (7, 1), network size 9 and two work zones (9, 2), 11 network size and three work zones (11, 3).
2. *Work zone size*: network size 7 and 25% capacity reduction, network size 9 and 50% capacity reduction, and network size 11 and 75% capacity reduction.
3. *Sub-area size*: one work zone and 25% capacity reduction, two work zones and 50% capacity reduction, and three work zones and 75% capacity reduction.

Figure 18 (a) through (c) show Guadalupe Street, 7th Street, and West MLK's chosen links in the network size order 7 with one work zone. This study chose four links for the boxplots at each location with these requirements,

1. The links cannot be the work zones themselves;
2. There is at least one link between the chosen links;
3. The number of the links are different when work zone size or sub-area size changes. The chosen links should be included in the all chosen networks at each location.

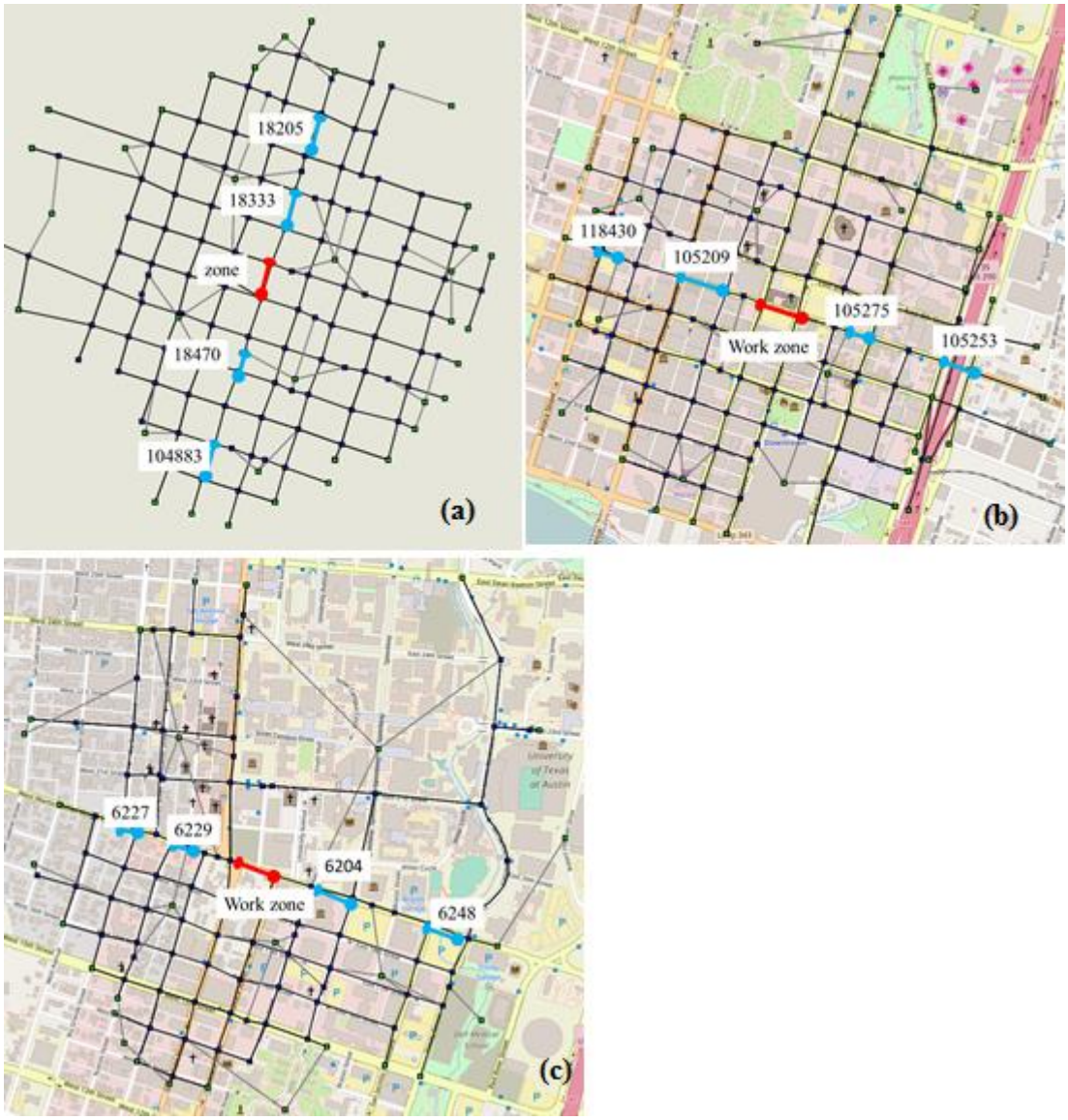


Figure 18 The chosen links in blue for the boxplots (a) Guadalupe Street, (b) 7th Street, and (c) West MLK. The work zone is shown in red.

Effect of Capacity Reduction on RMSE Values

As capacity reduction increases, the impacted scenario RMSE value increases. More lanes closed results in reduced link capacity, and travel times increase if the traffic flow does not change after the closure. To decrease travel times, vehicles reroute which leads to increased RMSE in the impacted scenario. Thus, this study assumes that when there is more capacity reduction, there will be larger RMSE values in the impacted scenario.

Guadalupe Street Network (7, 1)

Figure 19 shows the boxplots for RMSE values and capacity reduction for the four links (identified by ID Number) selected in the Guadalupe Street network. Figure 19 shows an increasing trend when the capacity reduction increases, except for link ID 104883. From the geometry network, link 104883 is located at the end where vehicles are leaving the work zone and is the most far away from the work zone. Link 18470 is the second less influenced by the capacity reduction. Its location is also at the end of the work zone, but closer to the work zone than link 104883. Conversely, link 18333 and link 18205 experience a larger impact and both are entering the work zone. Link 18333 has the largest impact and is closest to the work zone. Therefore, this study assumes that either the target link location (upstream or downstream) or the target link's distance to the work zone affect the RMSE values in the impacted scenario.

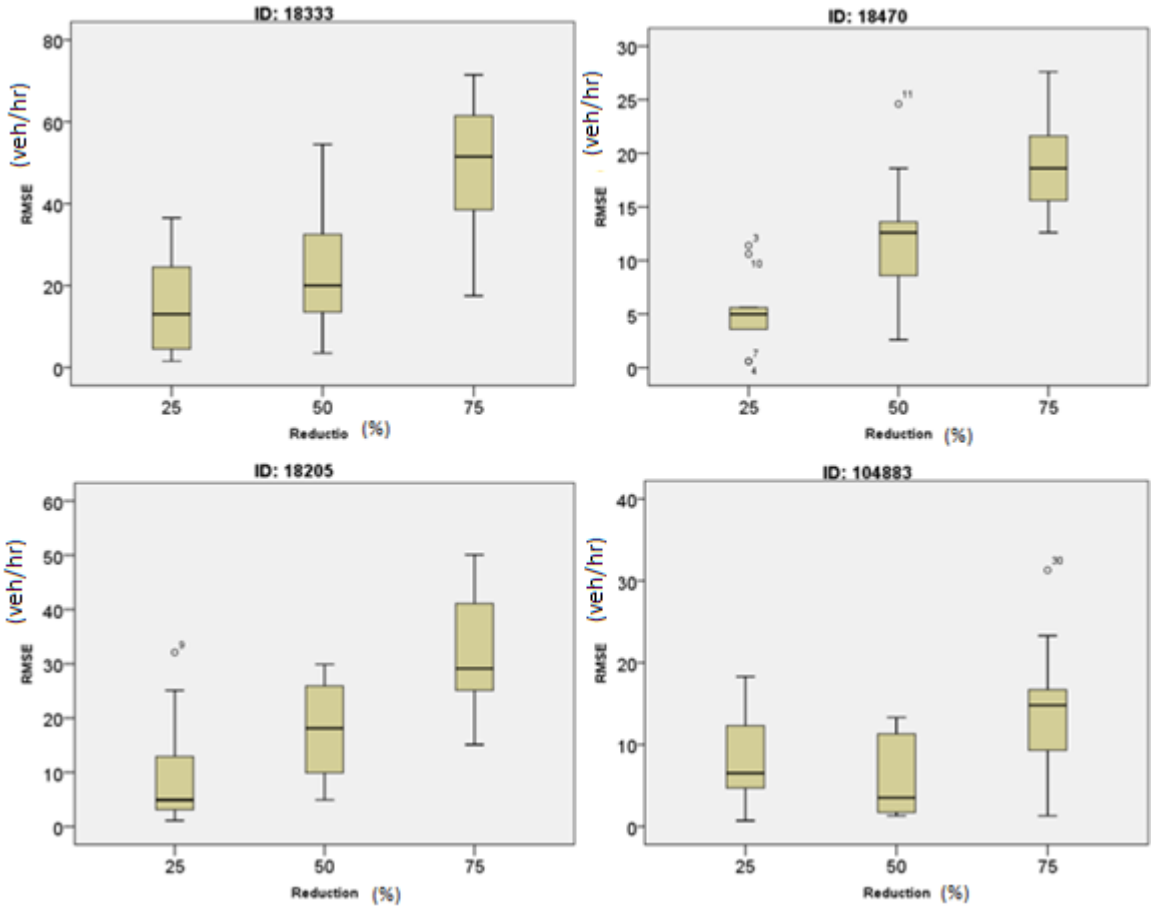


Figure 19 Boxplots for four selected links in the Guadalupe Street Network with a (7, 1) configuration undergoing capacity reduction. The upper set are close to the work

zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

7th Street Network (7, 1)

Figure 20 shows an ascending trend as capacity reduction increases. This result is similar to Guadalupe (7, 1). Also, the RMSE values increase much more rapidly going from 50% to 75% capacity reduction than when going from 25% to 50%. The v/c ratio of the work zone is 0.2 in the 7th Street network, which is larger than the v/c ratio of the Guadalupe network (= 0.16). Thus, the v/c ratio at the work zone influences the impacted area.

In addition, Figure 20 verifies that the distance from the target link to the work zone influences the RMSE values from the impacted scenario. Link 105275 and link 105209 both are close to the work zone, and their RMSEs from the impacted scenario reach ~ 125 and 80, respectively, at 75% capacity reduction. Conversely, link 105253 (located further away from the work zone) and link 118430 have values less than 20 at 75% capacity reduction.

However, Figure 20 conflicts with the previous assumption that the upstream links are impacted more by a work zone from Guadalupe (7, 1). Link 105209 is the upstream link and link 105275 is the downstream link, but link 105275 experiences a larger impact at 75% reduction.

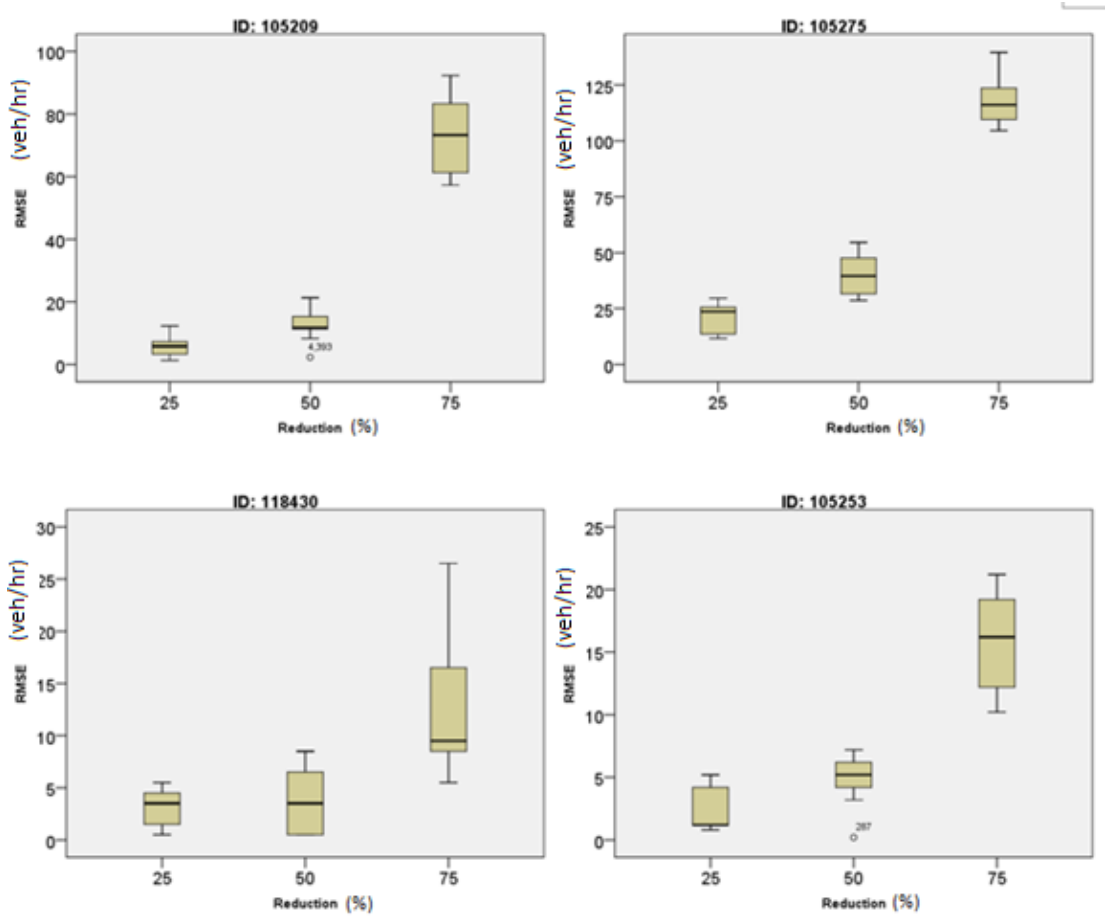


Figure 20 Boxplots for four selected links in the 7th Street Network with a (7, 1) configuration undergoing capacity reduction. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

West MLK Network (7, 1)

Figure 21 shows that the RMSEs from the impacted scenario in the West MLK network are different from RMSE values in 7th Street (7, 1) network. Link 6204 and link 6229 have an ascending trend when the reduction increases, similar to 7th Street. Conversely, link 6248's impact RMSE is identical at 25% and 50% capacity reduction, and link 6227's even decreases when going from 25% to 50% capacity reduction. Link 6248 and link 6247s' RMSEs are close to zero at 50% reduction, and link 6248's is zero at both 25% and 50% reduction. These links, which are the furthest away from the work zone at the upstream and downstream end, are affected by the distance to the work zone.

In addition, the closest links to the work zone, link 6204 and link 6229, have RMSE values that are twice as much as Guadalupe's RMSE values. When capacity reduction is 75%, link 6204 is around 120 vehicles per hour and link 6229 is around 80 vehicles per hour in West MLK; link 18333 is around **55** vehicles per hour and link 18470 in is around 20 vehicles per hour on Guadalupe street. MLK's work zone v/c ratio is 0.35, which is around twice as much as the Guadalupe work zone v/c ratio of 0.16. Thus verifying that the work zone's v/c in the base scenario influences the RMSEs calculated from the impacted scenario.

Link 6204 and link 6229 are both close to the work zone, but the traffic impacts are different with 50% reduction. Link 6204 is affected when the capacity reduction increases because the RMSE increases going from 25% to 50% reduction. Conversely, link 6229 is not affected from 25% to 50% capacity reduction because it has similar RMSE values in both cases. In addition, the RMSE at 75% capacity reduction for link 6204 is almost 50% higher than for link 6229. Link 6204 is close to the third work zone, upstream to the first work zone; link 6229 is close to the second work zone, downstream to the first work zone.

In the base scenario, the v/c ratios for the work zones are 0.35, 0.11, and 0.26 at the first, second (downstream to the first work zone), and third (upstream to the first work zone) work zone in the 7th Street network. Around 1/3 of the traffic flow (0.11/0.35) leaves West MLK street from the first work zone to the downstream link, and from Figure 10 in Chapter 3, around 2/3 of the traffic flow enters Guadalupe Street. It is intuitive that if less traffic flow stays on the same street, the traffic impact is also less if the capacity is reduced in the upstream link. Thus, from link 6204 and link 6229, it is reasonable to assume that the target link and the work zone's flow ratio or flow difference might influence the RMSE in the impacted scenario because link 6204's v/c ratio is similar to 0.26 (≈ to third work zone), and link 6229 is similar to 0.11 (≈ to second work zone).

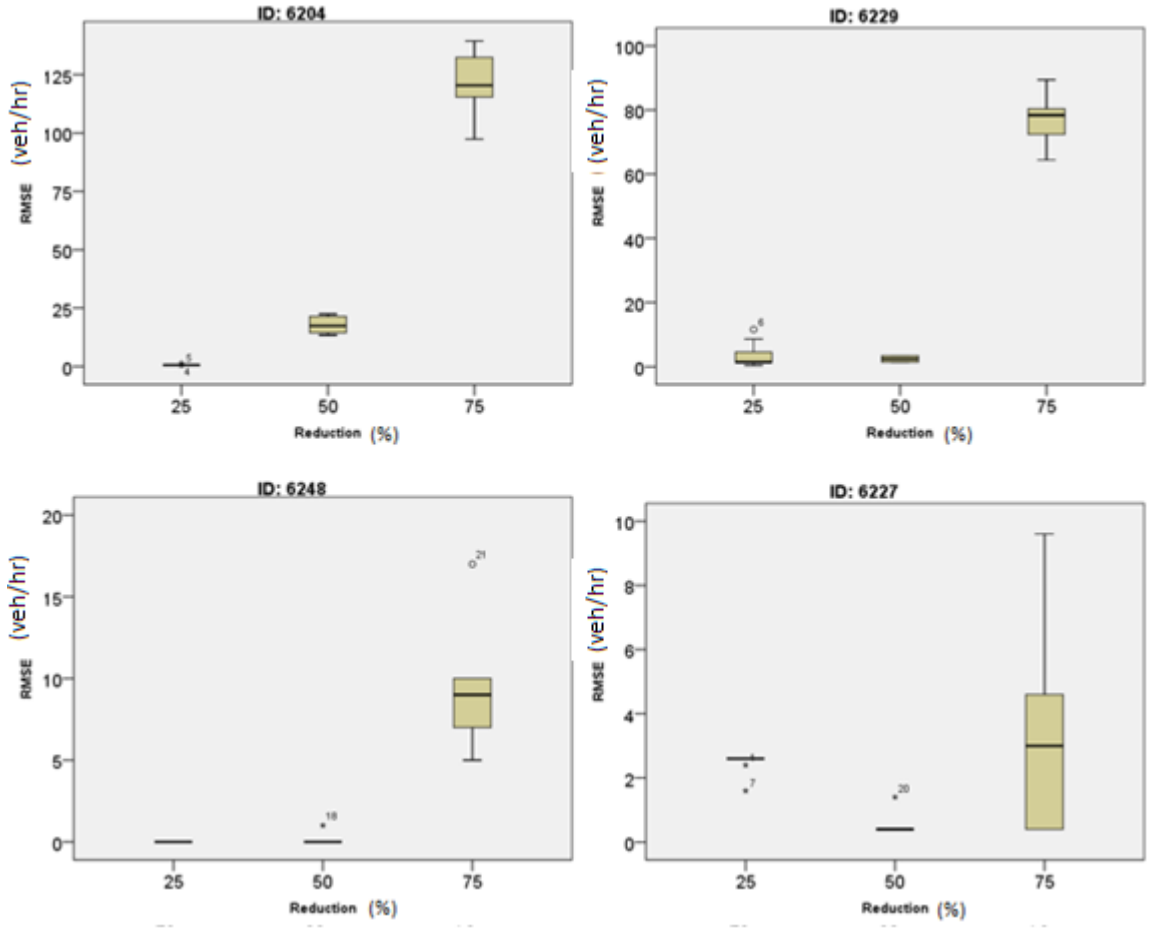


Figure 21 Boxplots for four selected links in the West MLK Network with a (7, 1) configuration undergoing capacity reduction. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

Guadalupe Street Network (9, 2)

Figure 22 shows linear ascending trend as capacity reduction increases when adding in an additional work zone at the first work zone's downstream link and increasing the sub-area size by one order size. The impact RMSEs at the upstream links, link 18333 and link 18205, are similar to Guadalupe (7, 1). Conversely, link 18470 and link 104883 RMSE values are ~60 at 75% reduction, which is ~ 20 higher than in Guadalupe (7, 1). This result verifies that the impact RMSEs are influenced by the distance between the work zone and the target link when the new work zone is close to the downstream links.

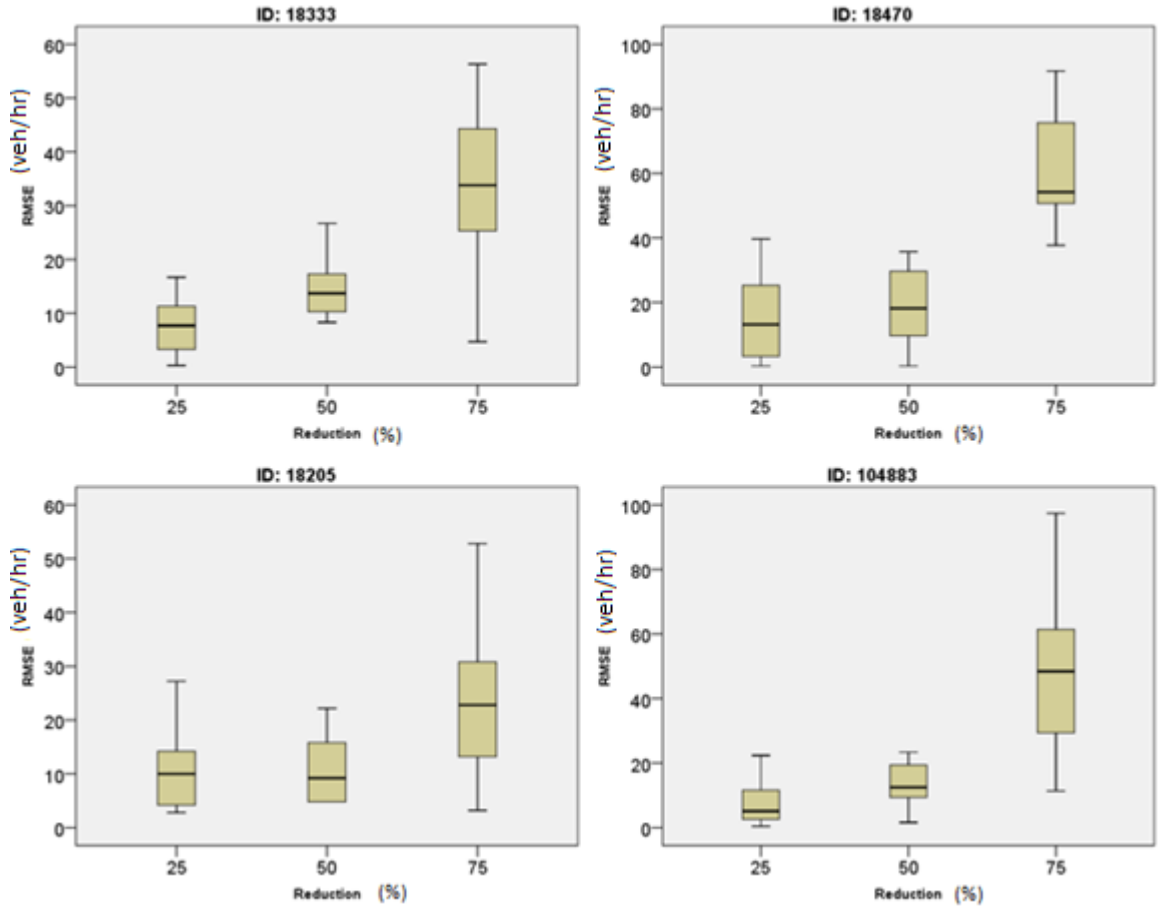


Figure 22 Boxplots for four selected links in the Guadalupe Network with a (9, 2) configuration undergoing capacity reduction. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

7th Street Network (9, 2)

Figure 23 shows an increasing trend as capacity reduction increases. Link 118430 experiences the lowest impact from the work zone. Also at the downstream links, the impact RMSEs at link 105275 and link 105253 are around 80 vehicles per hour larger than the 7th Street (7, 1) network because the new work zone is close to the downstream links.

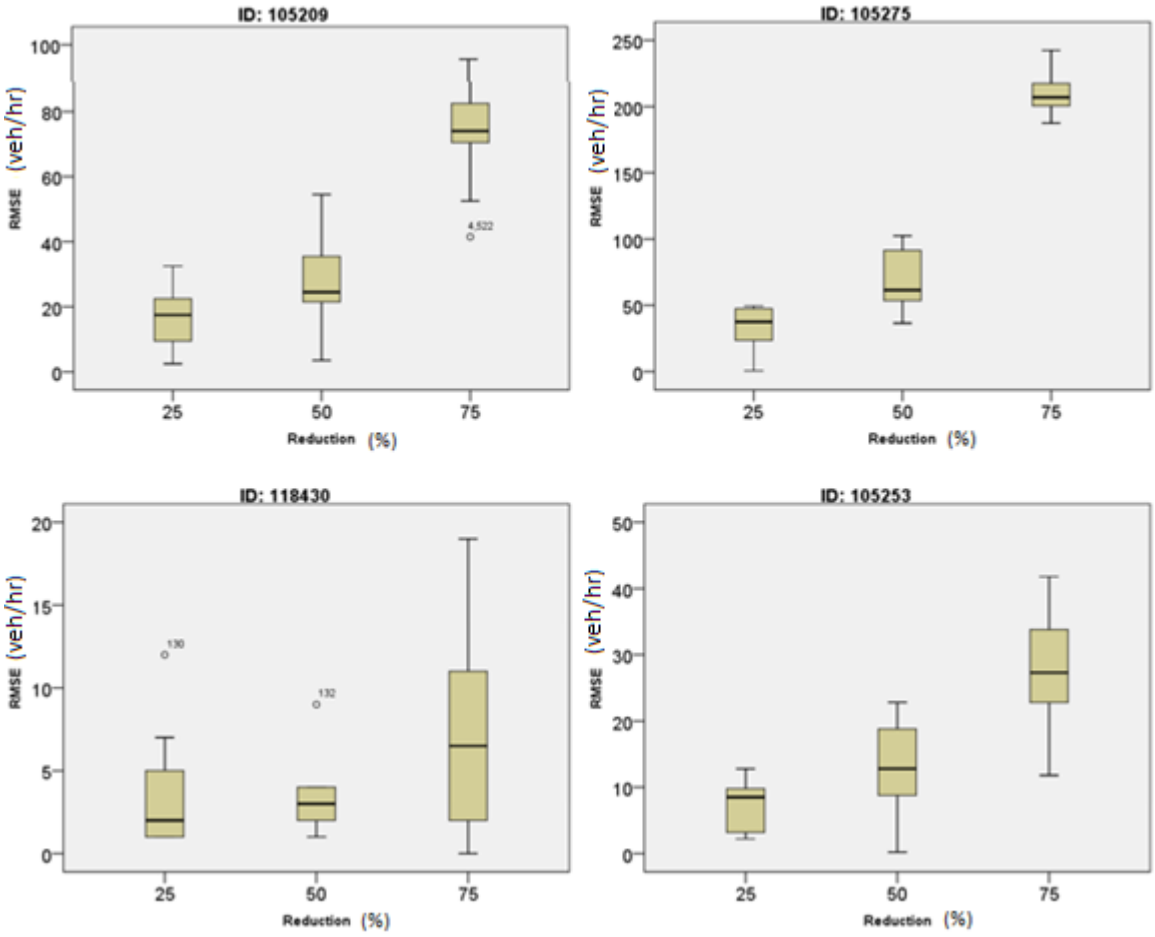


Figure 23 Boxplots for four selected links in the 7th Street Network with a (9, 2) configuration undergoing capacity reduction. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

West MLK Street Network (9, 2)

Figure 24 shows that the RMSEs from the impacted scenario increase as the capacity reduction increases. The closest links to the work zone, link 6227 and link 6229, were impacted more than the furthest away links, link 6248 and link 6227. However, the new work zone was close to the downstream target links, but these links were impacted less than the downstream links in the Guadalupe (9,2) Network and the 7th Street (9, 2) Network, which merely increased around 20 vehicles per hour in comparison to West MLK (7, 1). The new work zone's v/c was only around 0.11 in the base scenario, and the travel time did not increase and the vehicles did not reroute when the capacity was reduced. Thus, RMSEs from link 6229 and link 6227 proved that the work zone's v/c in the base scenario affects the target link. The target link was not influenced by the work zone's low v/c even though the target link was located close to the work zone.

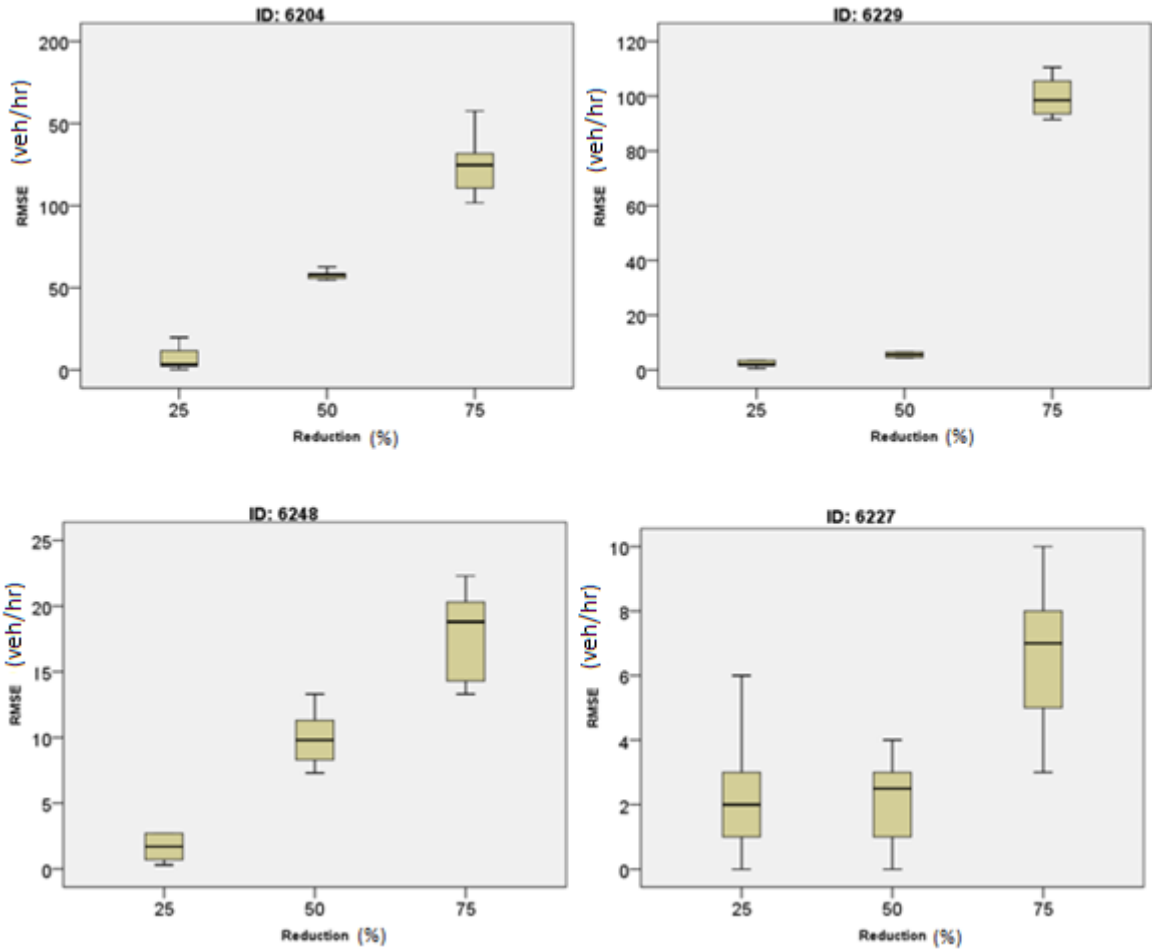


Figure 24 Capacity reduction boxplots for four selected links in the West MLK Network with a (9, 2) configuration. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

Guadalupe Street Network (11, 3)

Here, one work zone link connected to the first work zone's upstream link was added, which increased the network size by one order. Figure 25 shows that four links have ascending trends when capacity reduction increases. Since the new work zone was close to the upstream link, link 18333 (located at the upstream end and the closest to the work zone) was impacted most at 75% capacity reduction. Also, the other upstream link, link 18205, RMSE value increased by 55 vehicles per hour going from the (9, 2) configuration to the (11, 3) configuration at 75% reduction.

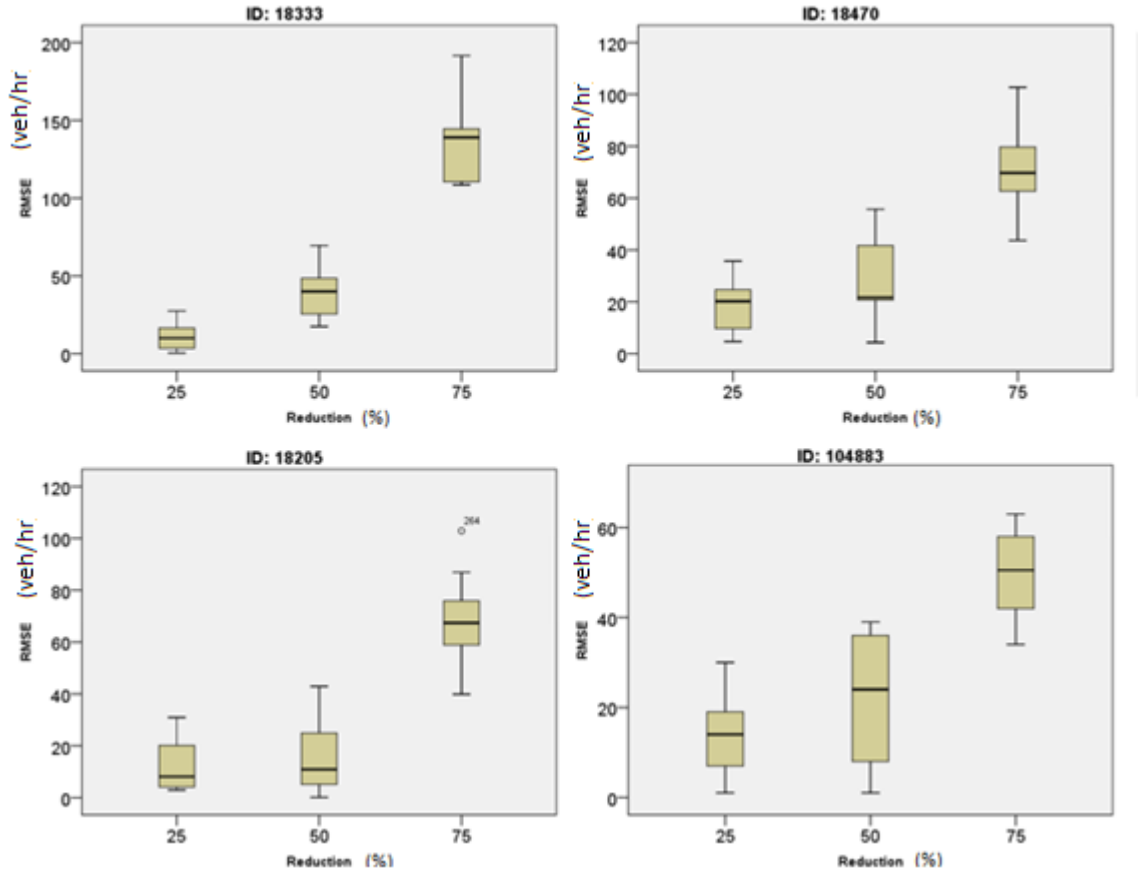


Figure 25 Boxplots for four selected links in the Guadalupe Network with a (11, 3) configuration undergoing capacity reduction. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

7th Street Network (11, 3)

From Figure 26, the links show an ascending trend as capacity reduction increases, except for link 118430. Link 118430's impact RMSE values equal roughly 10 vehicles per hour in the (7, 1), (9, 2), and (11, 3) configurations, implying that the link was not impacted by the work zone in these networks. Also, link 105209's RMSEs values are similar across the network sizes. There is no significant difference between the number of the work zones and the sub-area size at link 105209. While the number of vehicles decreased at the upstream links when there was capacity reduction, there were no additional vehicles that rerouted after adding in a third work zone.

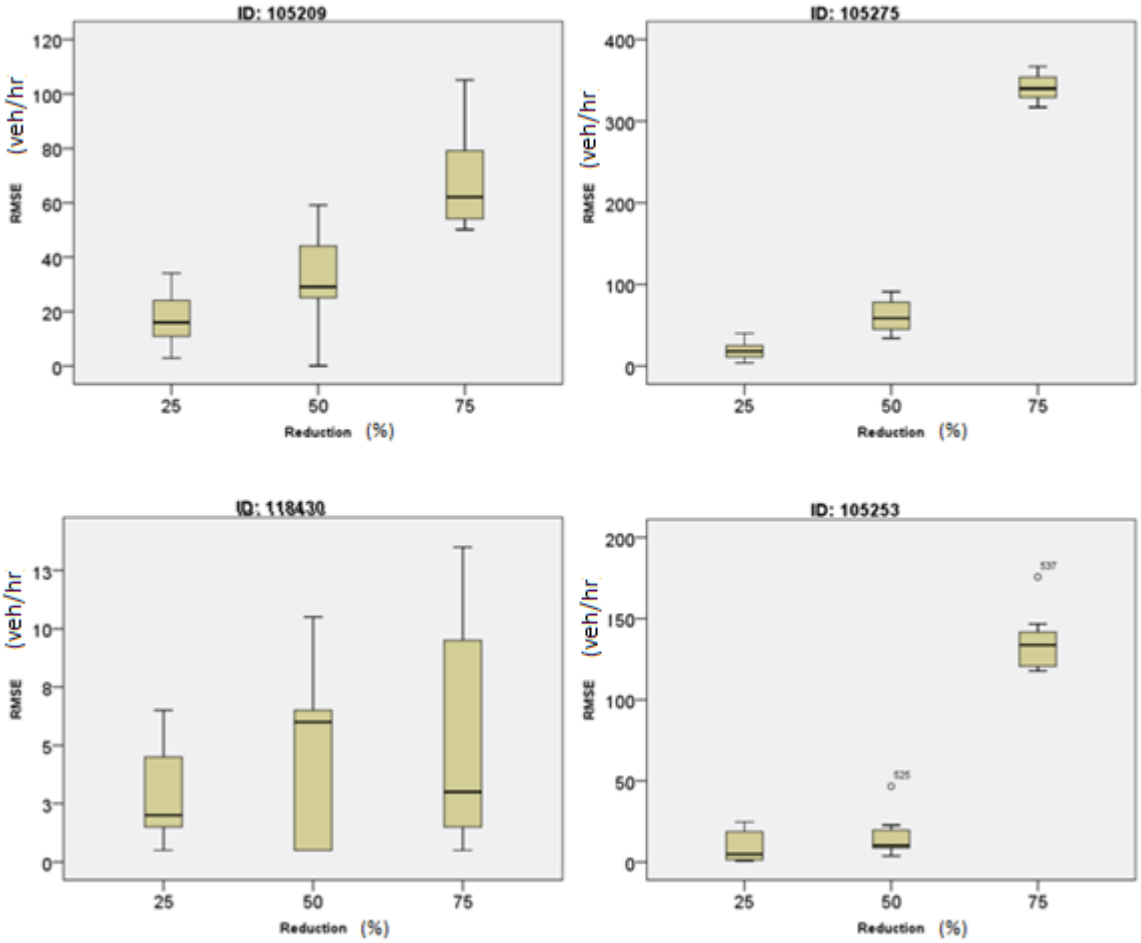


Figure 26 Boxplots for four selected links in the 7th Street Network with a (11, 3) configuration undergoing capacity reduction. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

West MLK Street Network (11, 3)

Figure 27 displays the impact RMSE values after adding an additional work zone connected to the first work zone's upstream link. Only link 6227 impact RMSE values did not increase when the capacity reduction increased. Link 6227 is located downstream and faraway from the work zone, so it is reasonable that this target link might not receive any impact from adding an additional work zone.

The upstream target links, link 6204 and link 6248 increase by roughly 60 vehicles per hour from (7, 1) to (11, 3) as a result of adding an extra work zone upstream. Conversely, the downstream target links, link 6229 and link 6227, did not change much between (9, 2) and (11, 3). This result verifies that the number of the work zones did not affect the impact RMSE values, but that the distance between the work zone and the target link does.

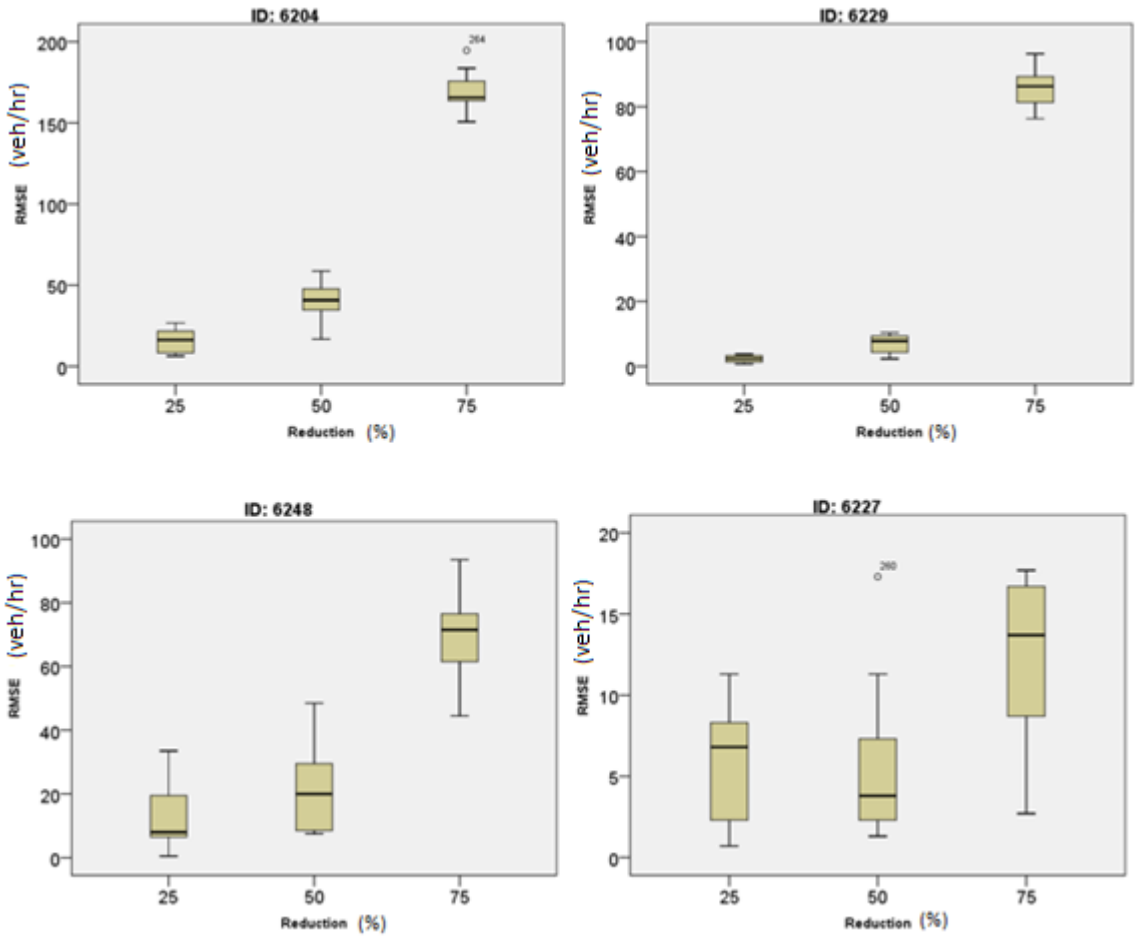


Figure 27 Boxplots for four selected links in the West MLK Network with a (11, 3) configuration undergoing capacity reduction. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

Capacity Reduction Summary

This study boxplots show increasing RMSE values with increasing capacity reduction for the Guadalupe, 7th Street, and West MLK's (7, 1), (9, 2), and (11, 3) networks. There are several findings, including:

1. The work zones' capacity reduction affects the target links.
2. The distance between the work zone and the target link influences RMSE values.
3. The number of the work zones does not necessarily affect RMSE values.

Effect of Work Zone Size on RMSE Values

Networks with (7, 25%) Configuration

Figure 28 through 30 show that there is not a consistent trend between work zone size and RMSE values, and that the RMSE values are stable or have no trend between work zone sizes in the three networks, Guadalupe Street, 7th Street, and West MLK, with a (7, 25%) configuration In Guadalupe Street Network, the RMSE differences are less than five vehicles from one to three work zones. In 7th Street Network, Link 105209 and link 105253 seems to have ascending trend when the number of impacted links increases, but the other two links' RMSEs values decrease when the number of impacted links increases from two to three. In West MLK Street Network, the average RMSEs are almost equal from one to three work zones across link 6229, 6248, and 6227. This result might verify that the number of the work zones does not affect the impact RMSEs.

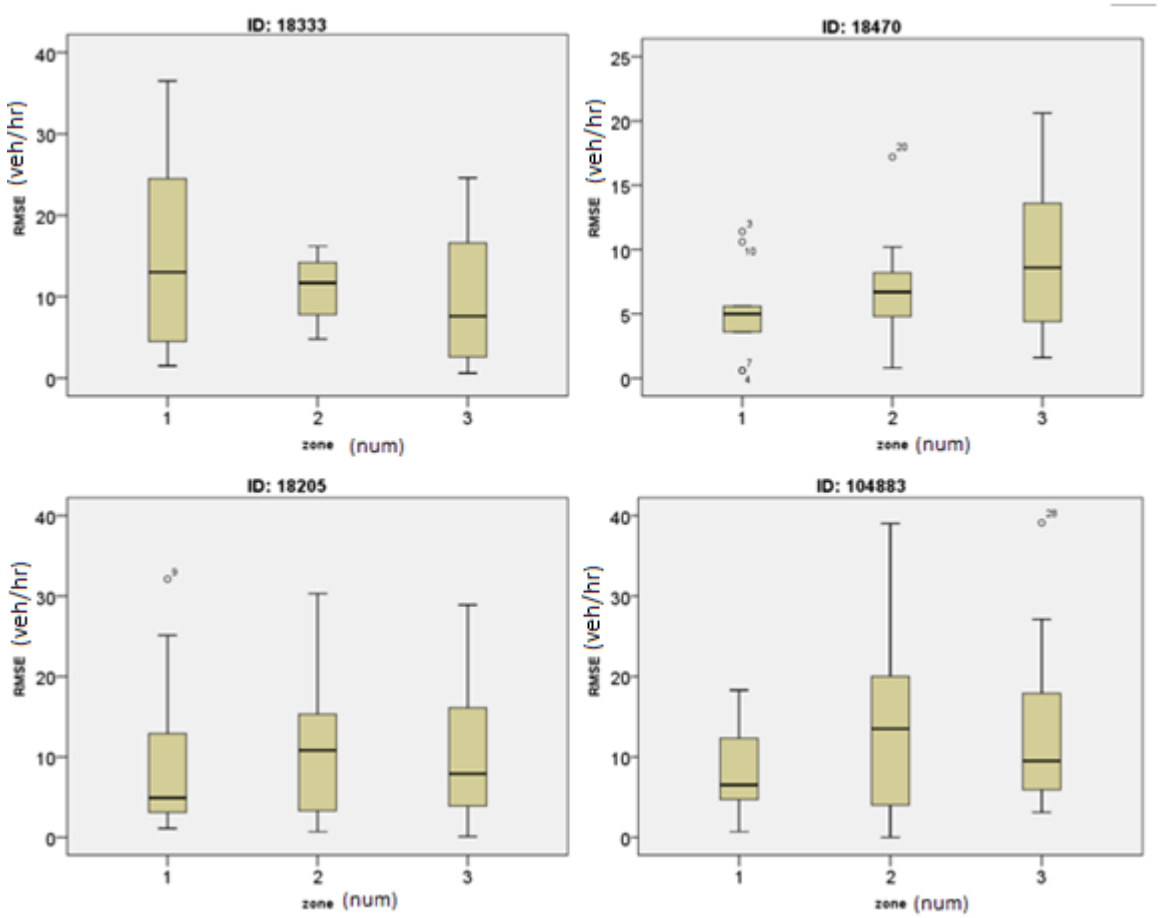


Figure 28 Boxplots for four selected links in the Guadalupe Network with a (7, 25%) configuration with varying work-zone sizes. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

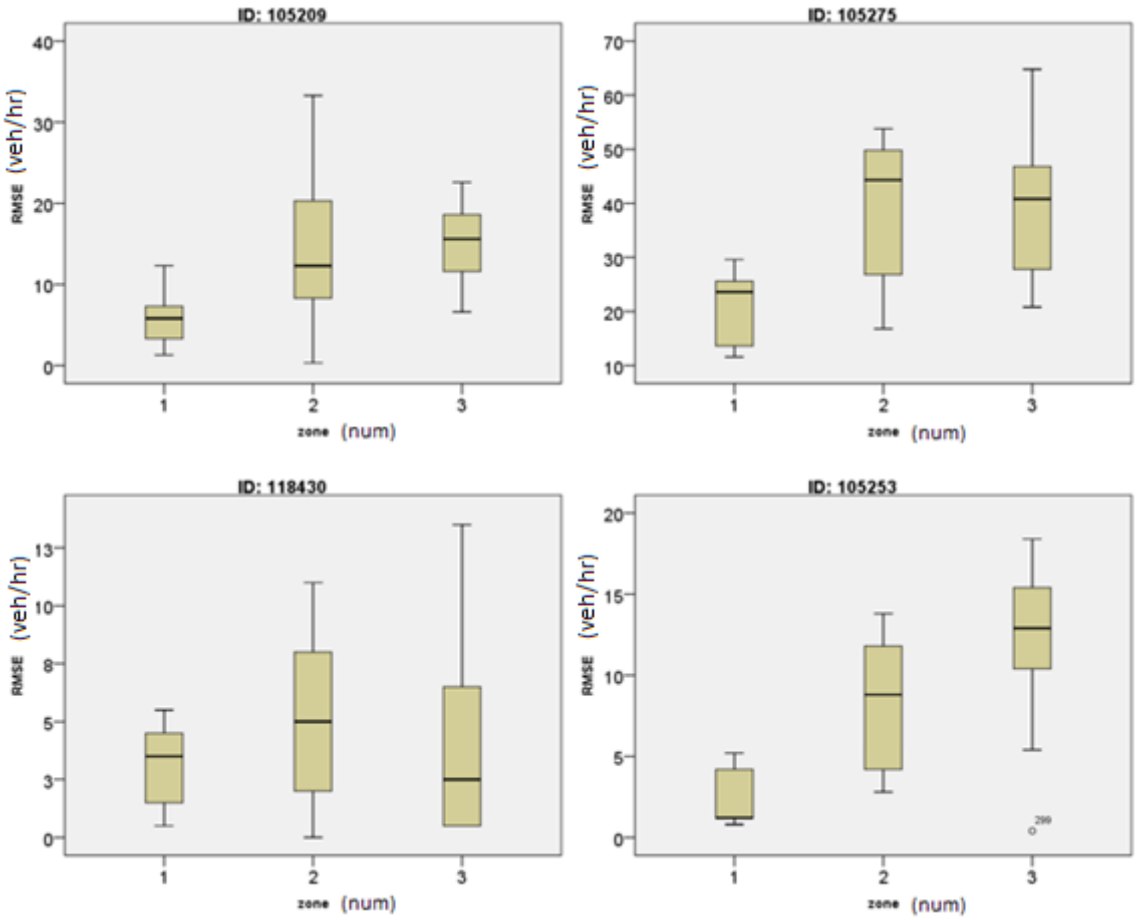


Figure 29 Boxplots for four selected links in the 7th Street Network with a (7, 25%) configuration with varying work-zone sizes. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

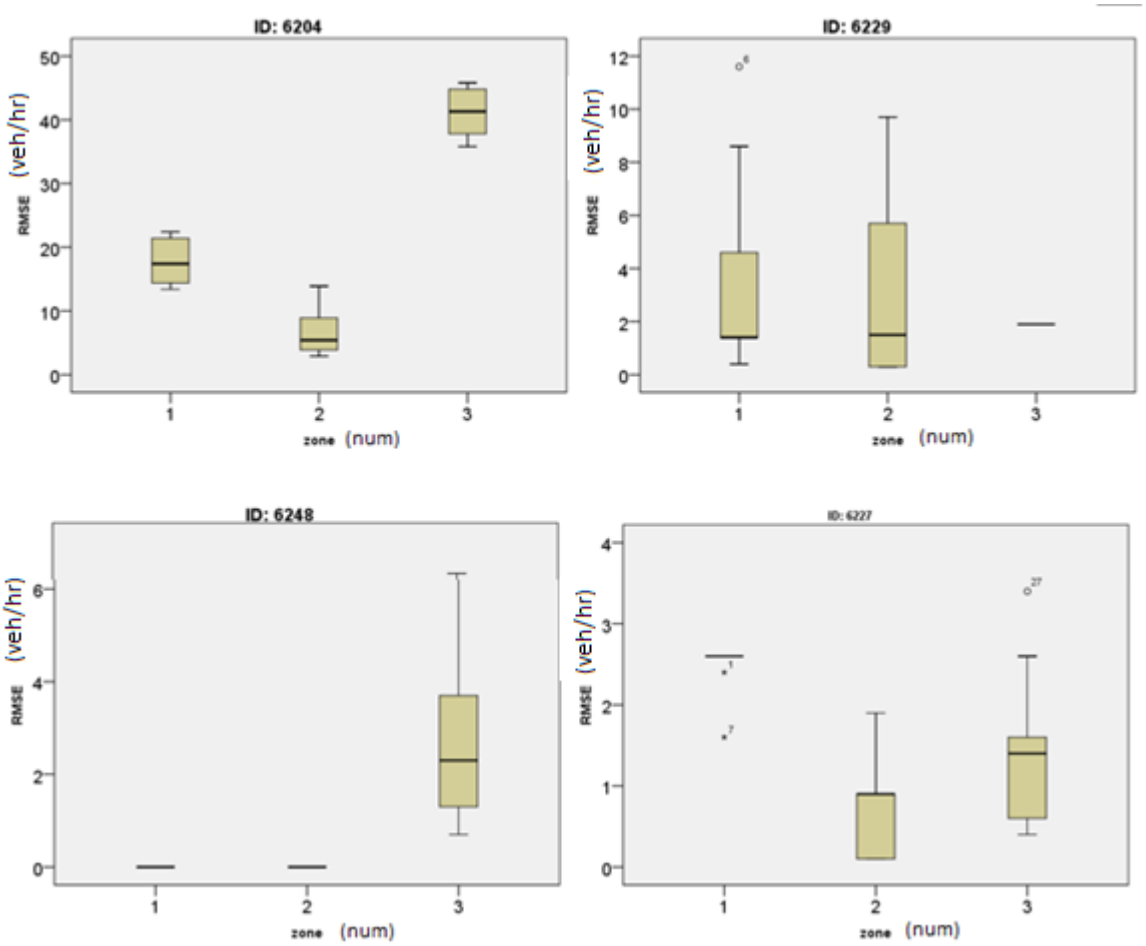


Figure 30 Boxplots for four selected links in the West MLK Network with a (7, 25%) configuration with varying work-zone sizes. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

Networks with (9, 50%) Configuration

Figure 31 through Figure 33also show that there is no consistent trend and that the impact RMSEs are similar between the work zone size and the impact RMSEs (9, 50 %) across three locations, Guadalupe, 7th, and West MLK. Conversely, Figure 31 shows the work zone location impact on the target links. For link 18333 and link 18205, the upstream target link RMSE values are similar at 25% and 50% capacity reduction, but increase more at 75% capacity reduction (Figure 31). The second work zone is located at the first work zone's downstream end, and the third work zone is at the upstream end (see Figure 10 in Chapter 3). Thus, the upstream target links are not influenced by two work zones, but by three work zones. In addition, link 18470's RMSE value also increases with two work zones, where the second work zone is close to link 18470. Thus, this figure verifies that the impact RMSE values are influenced by the distance to the work zone.

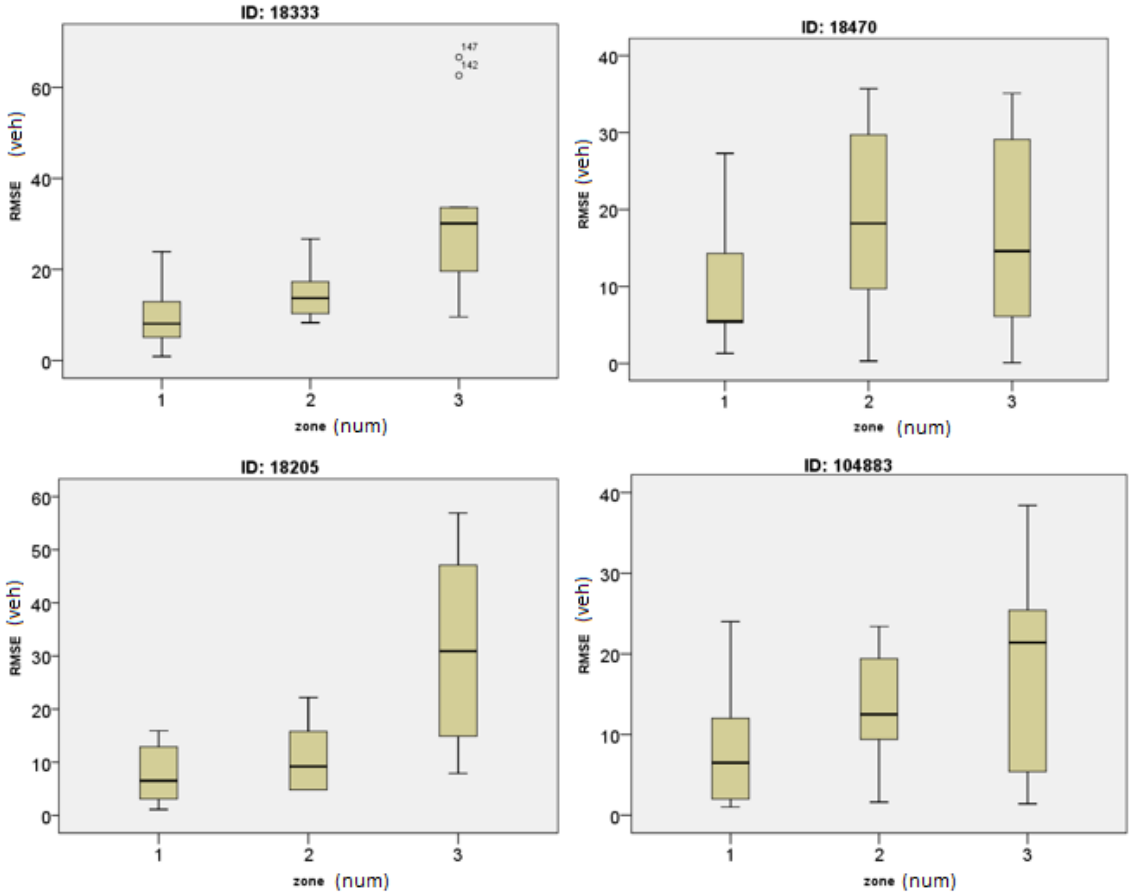


Figure 31 Boxplots for four selected links in the Guadalupe Street Network with a (9, 50%) configuration with varying work-zone sizes. .The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

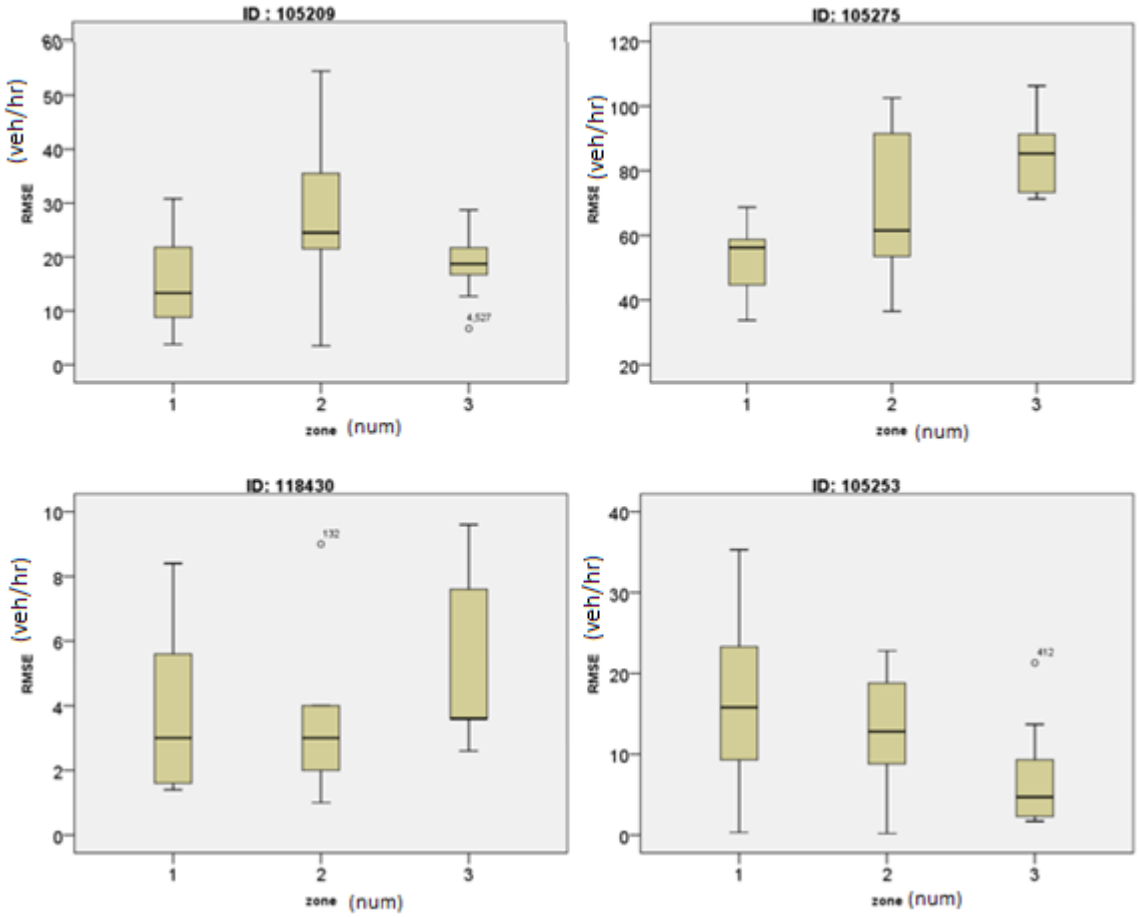


Figure 32 Boxplots for four selected links in the 7th Street Network with a (9, 50%) configuration with varying work-zone sizes. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

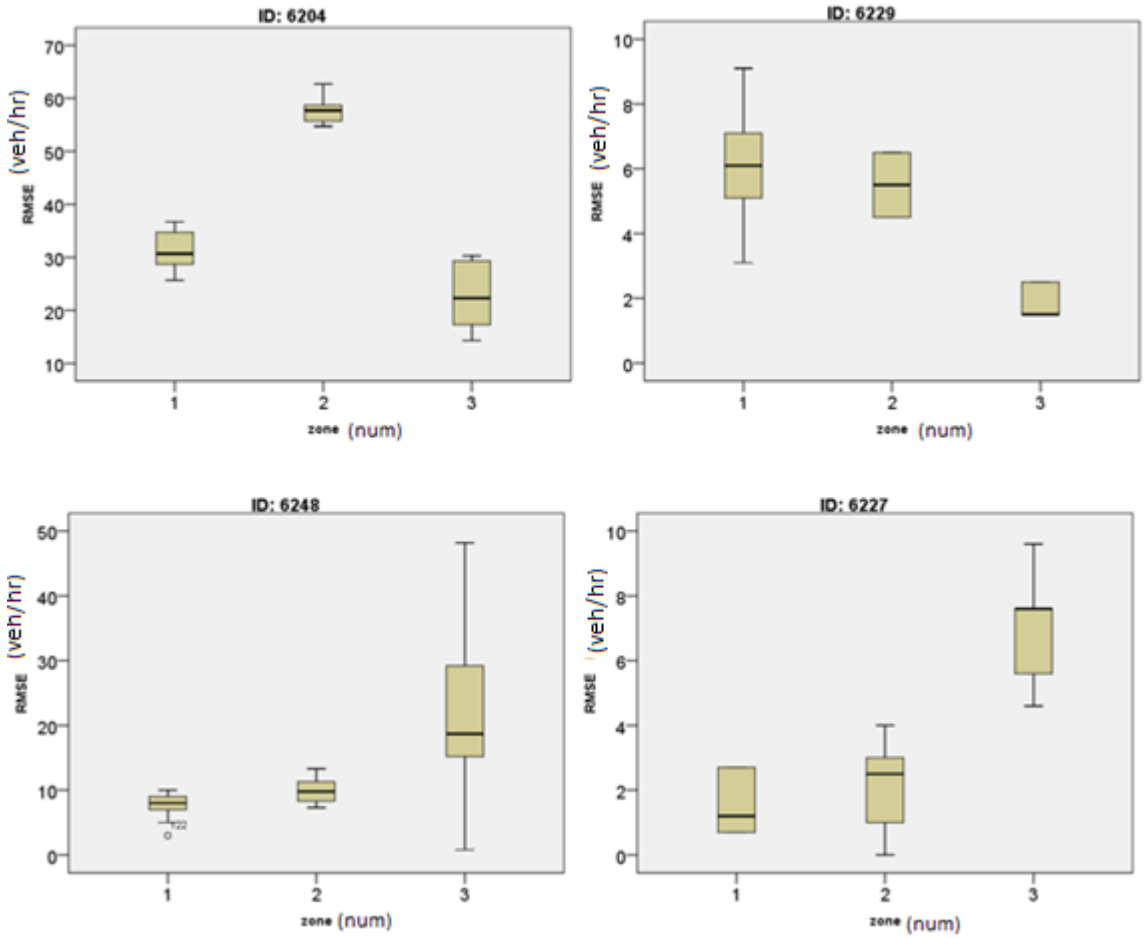


Figure 33 Boxplots for four selected links in the West MLK Network with a (9, 50%) configuration with varying work-zone sizes. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

Networks with (11, 75%) Configuration

Figure 34 through Figure 36 are boxplot RMSEs over the number of the work zones when the sub-area size is 11 and the work zone capacity reduction is 75%. From these figures, no consistent trend with varying work zone size can be determined from these target links. However, Figure 34 shows that the distance between the work zone and the target link affects the RMSE values in the impacted scenario. The second and the third work zones are located at the first work zones' downstream and upstream ends, respectively, and link 18470 and link 18333 are located at the downstream and upstream positions, respectively. Thus, link 18333 has similar impact from one and two work zones, and link 18470 has similar impact from two and three work zones.

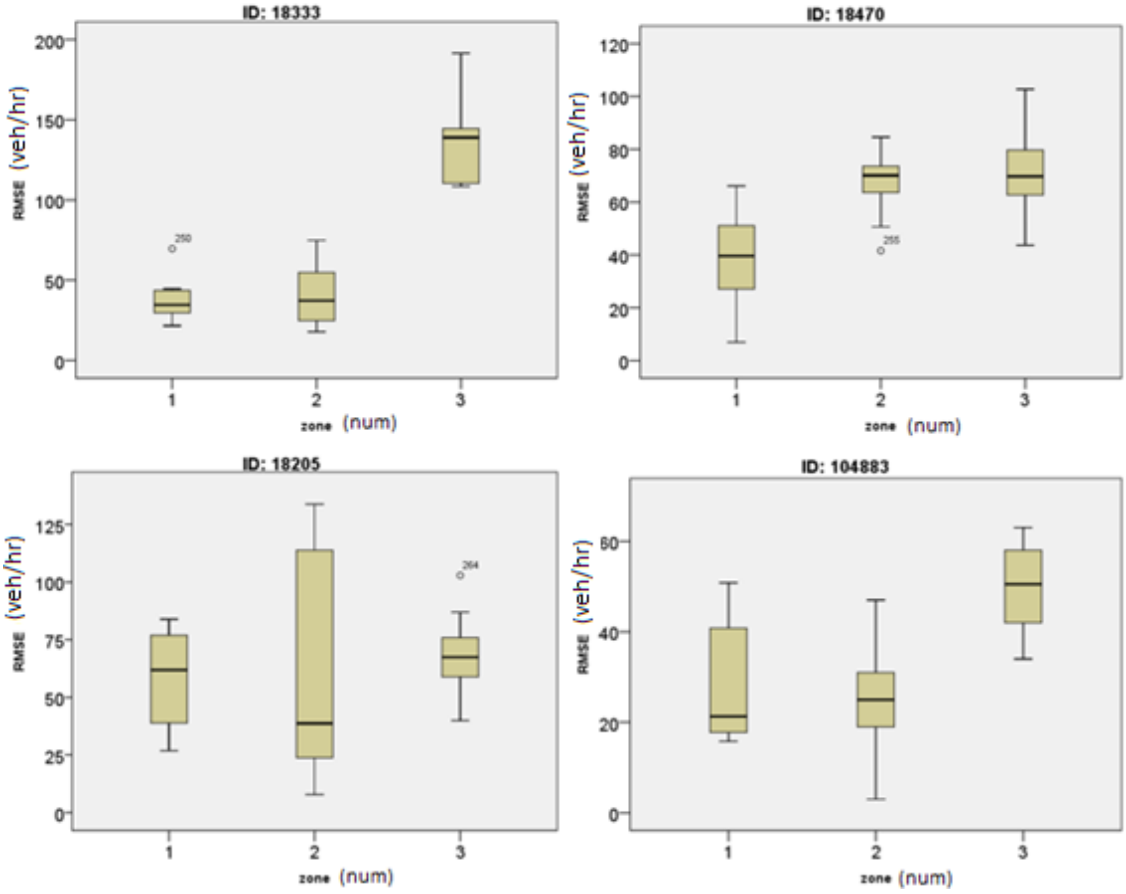


Figure 34 Boxplots for four selected links in the Guadalupe Network with a (11, 75%) configuration with varying work-zone sizes. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

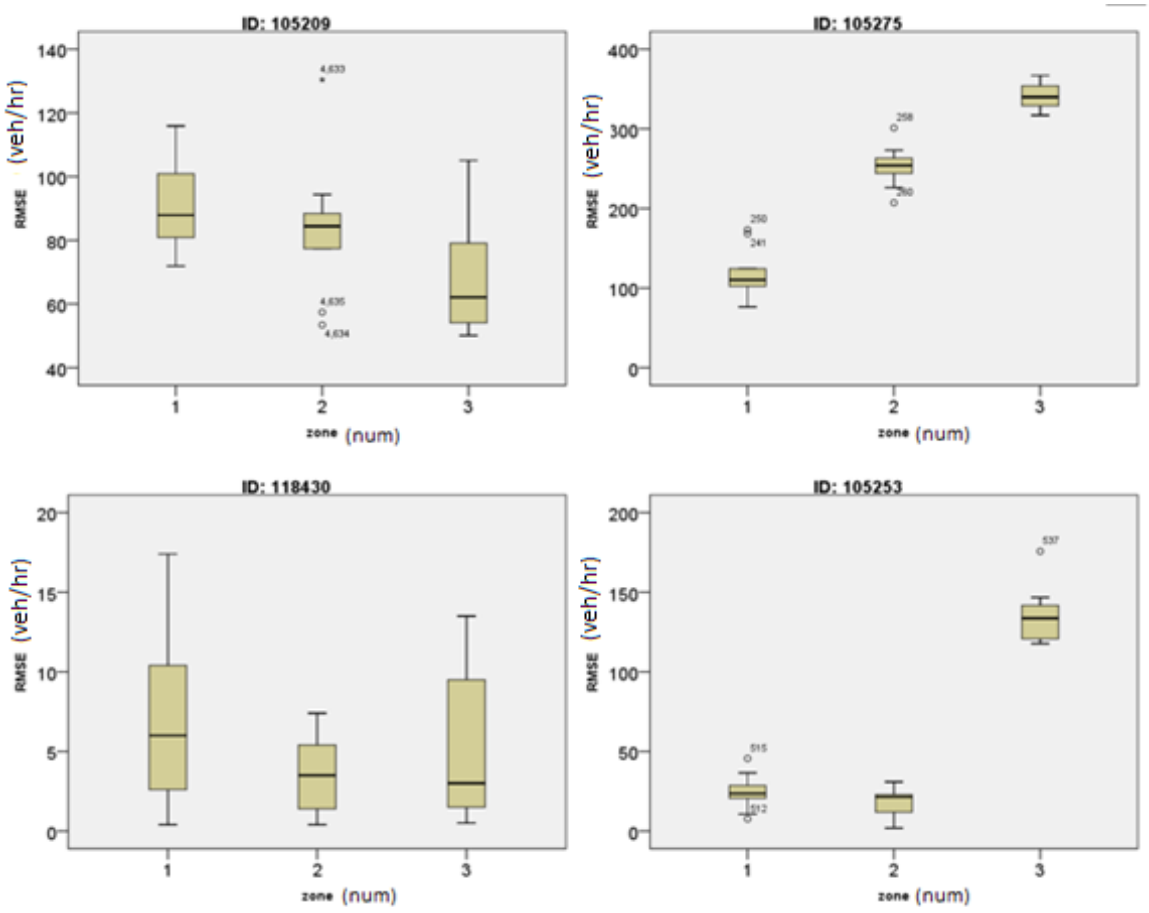


Figure 35 Boxplots for four selected links in the 7th Network with a (11, 75%) configuration with varying work-zone sizes. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

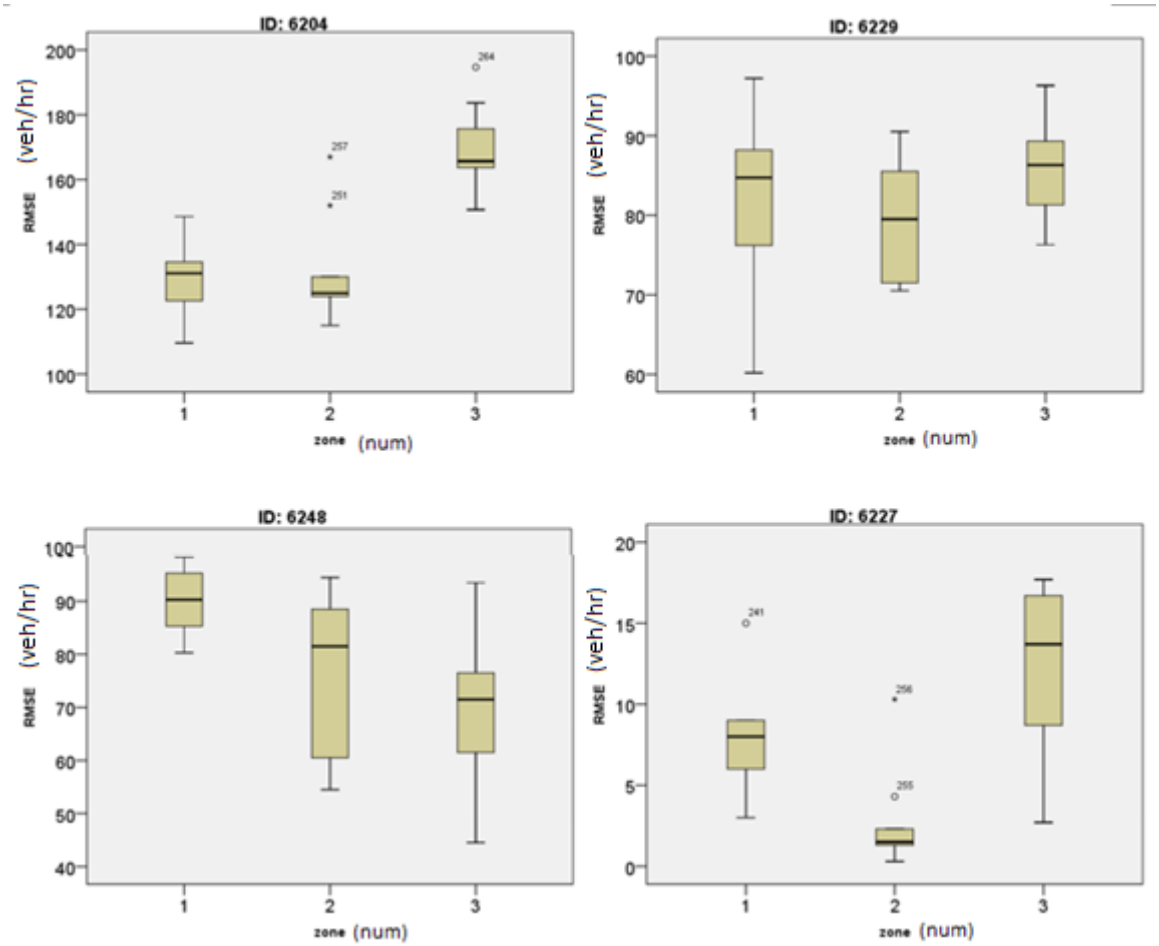


Figure 36 Boxplots for four selected links in the West MLK Network with a (11, 75%) configuration with varying work-zone sizes. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

Work Zone Size Summary

This study investigated the relationship between the number of work zones and the RMSEs values in these impacted scenarios. From the Capacity Reduction sectio*n,* this study found that the number of work zones does not affect the RMSE values, but the distance between the work zones and the target link does. Therefore, this study boxplotted the RMSE values with increasing number of work zones (from one to three work zones) across the following network configurations: (7, 25%), (9, 50%), and (11, 75%).

Figure 34 through 36 verify that increasing the number of work zones does not affect the impact RMSEs because there were no consistent patterns found in the target links. The random seeds in the simulation might cause the inconsistent pattern observed. Thus, this study will not consider work zone size as part of the impacted regression model.

Effect of Sub-Area Size on RMSE Values

For analyzing the relationship between the sub-area size and the impact RMSEs, this study boxplotted the RMSE values from subnetwork size 7, 9, and 11 across (1, 25%), (2, 50%), and (3, 75%) in Guadalupe Street, 7th Street, and West MLK networks. There are three assumptions for increasing the subnetwork size,

1. The RMSE values increase as the subnetwork size increases.

   For the objective of traffic assignment, all vehicles traveling from the same origin to the same destination will choose an identical short travel time path. More vehicle path choices occur as the sub-area enlarges. Thus, when there is a work zone in a large sub-area, vehicles have more chances to change their paths causing larger RMSE values.

2. The RMSE values decrease as the subnetwork size increases.

   The path travel time increases in a larger size sub-network, if the regional origins and destinations are outside of the sub-area. The work zone's effect decreases with a larger sub-area. Thus, vehicles do not reroute when subnetwork size increases.

3. There is no significant difference between the subnetwork sizes.

   The sub-area sizes do not affect vehicle rerouting, so the RMSEs do not have a consistent trend when the sub-area sizes increase.

Networks with (1, 25%) Configuration

Figure 37 through Figure 39 show that there is no consistent trend across the locations. Also, most of links show that the average impact RMSEs are similar when the sub-area size increases. It seems that the impact RMSEs are merely affected by the random seeds.

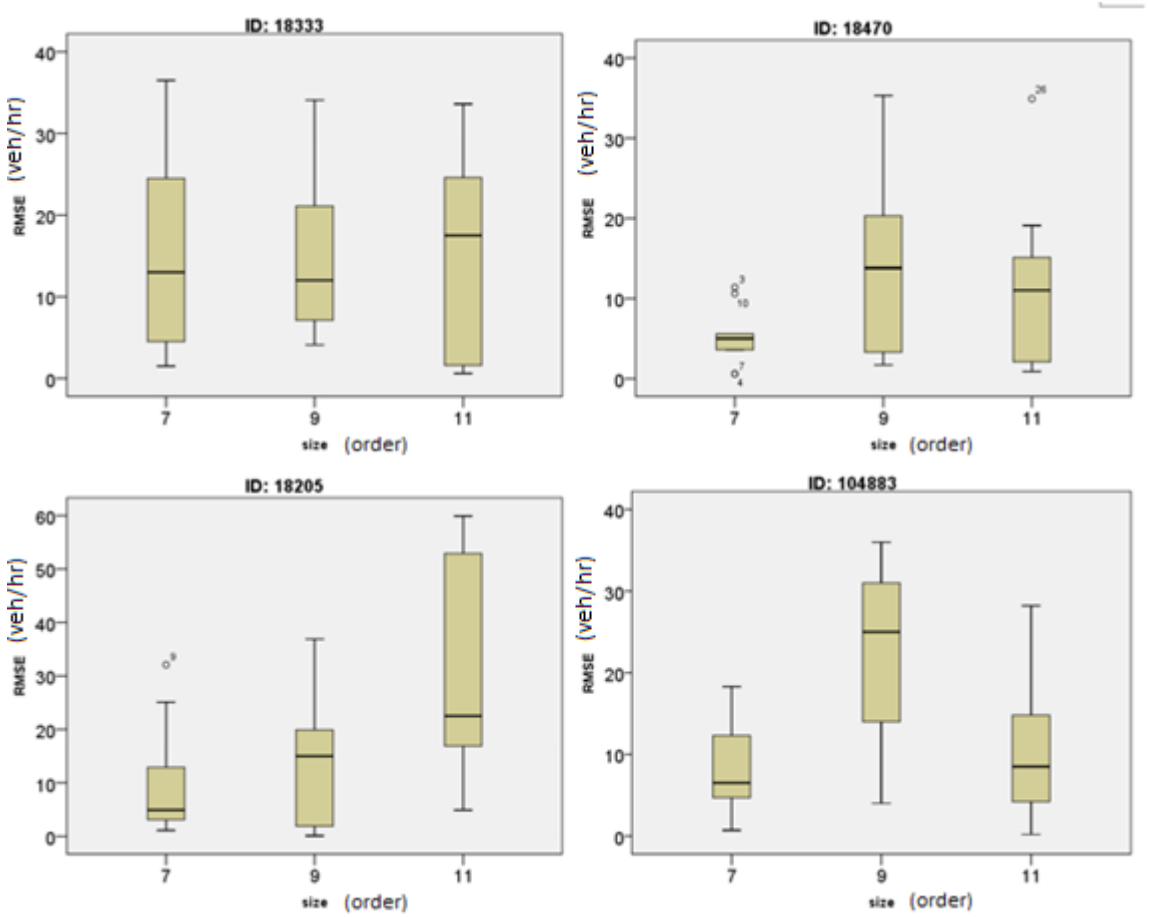


Figure 37 Boxplots for four selected links in the Guadalupe Street Network with a (1, 25%) configuration with varying sub-area size. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

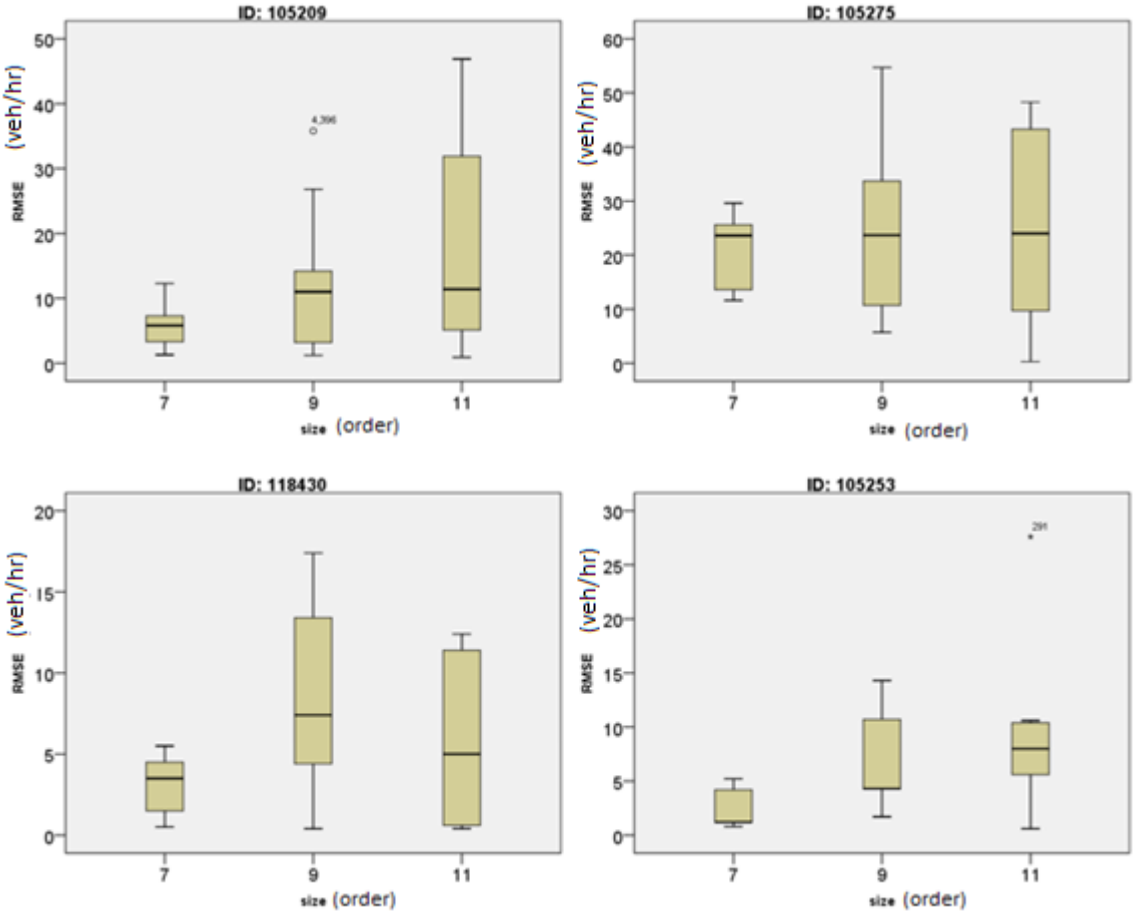


Figure 38 Boxplots for four selected links in the 7th Street Network with a (1, 25%) configuration with varying sub-area size. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

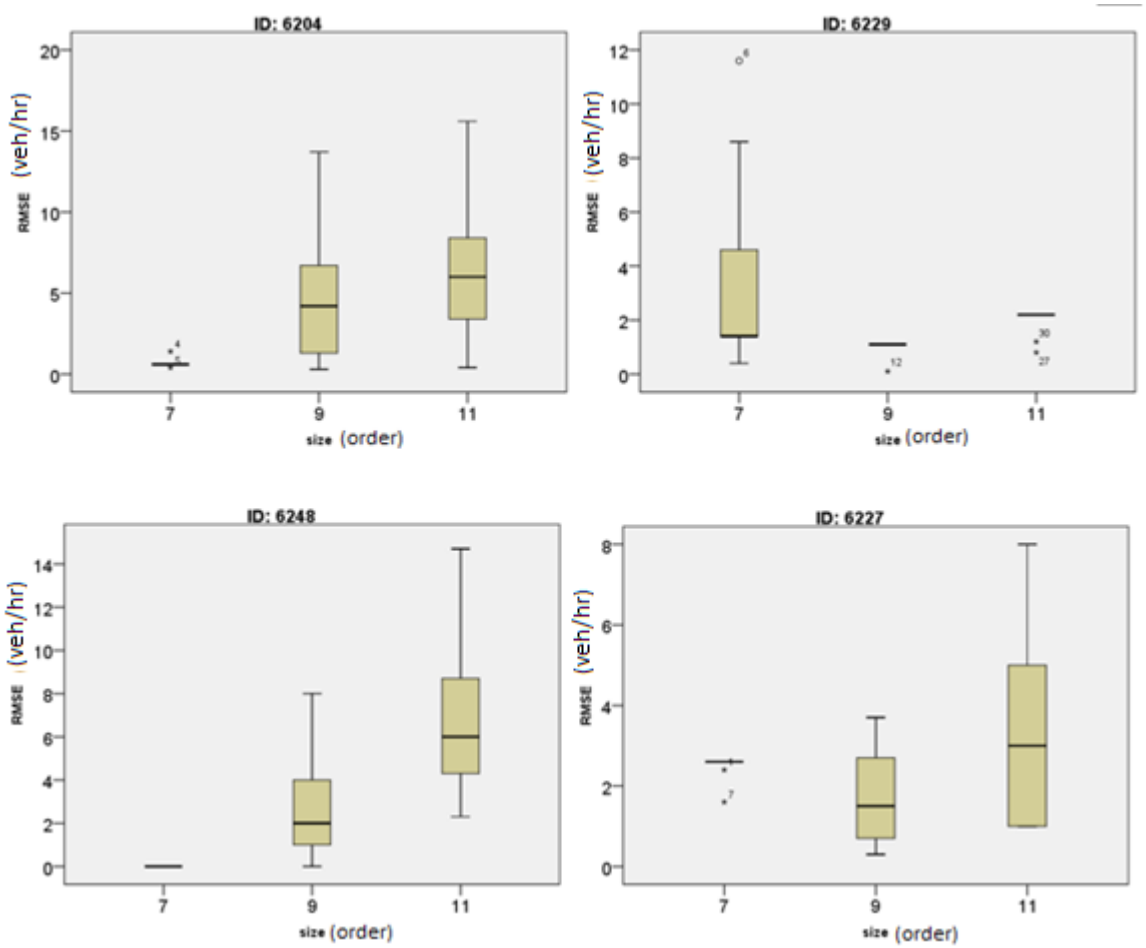


Figure 39 Boxplots for four selected links in the West MLK Network with a (1, 25%) configuration with varying sub-area size. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

Networks with (2, 50%) Configuration

Figure 40 through Figure 42again show an inconsistent pattern when the sub-area size increases across the three networks with two work zones and 50% capacity reduction. In addition, average RMSEs are also similar for the majority of the links. This result shows that the difference between the RMSEs can be attributed to the random seeds in the simulation rather than the sub-area size.

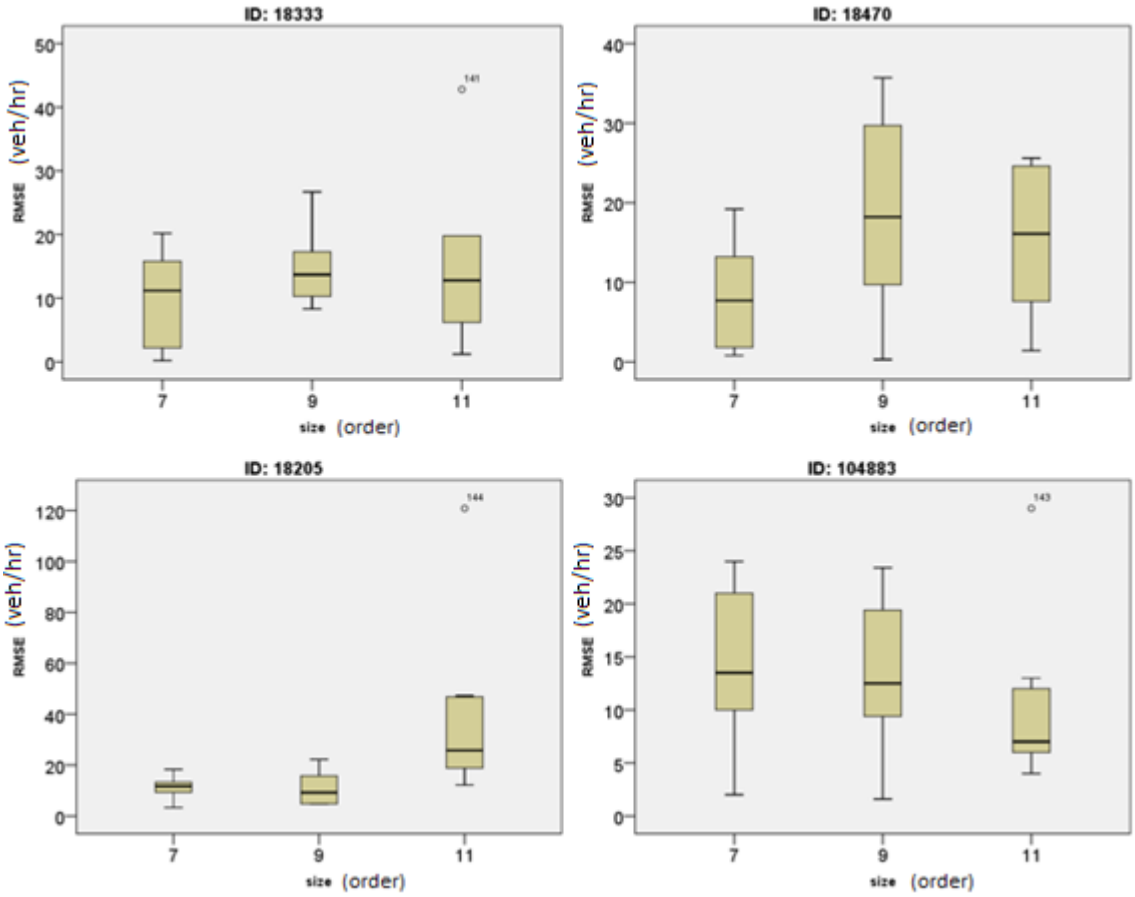


Figure 40 Boxplots for four selected links in the Guadalupe Street Network with a (2, 50%) configuration with varying sub-area size. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

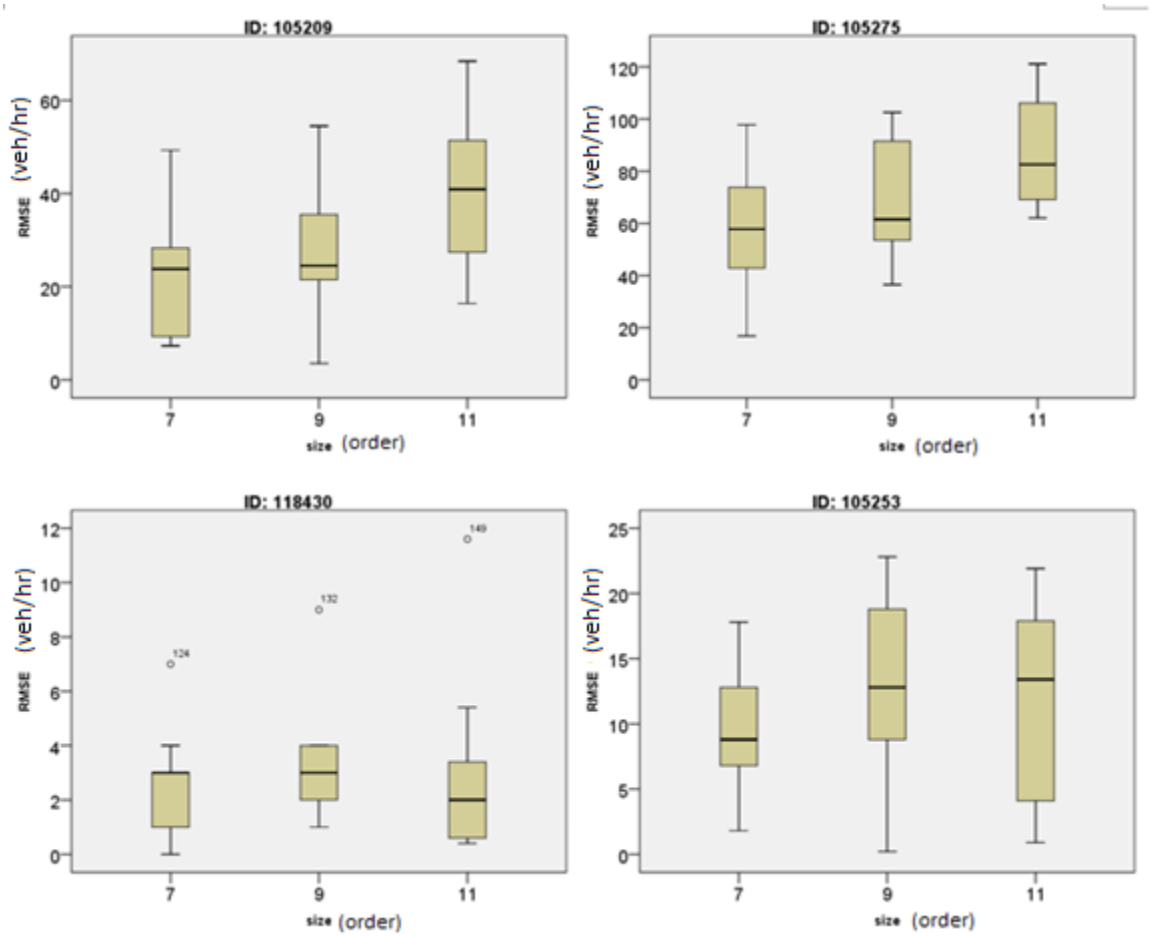


Figure 41 Boxplots for four selected links in the 7th Street Network with a (2, 50%) configuration with varying sub-area size. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

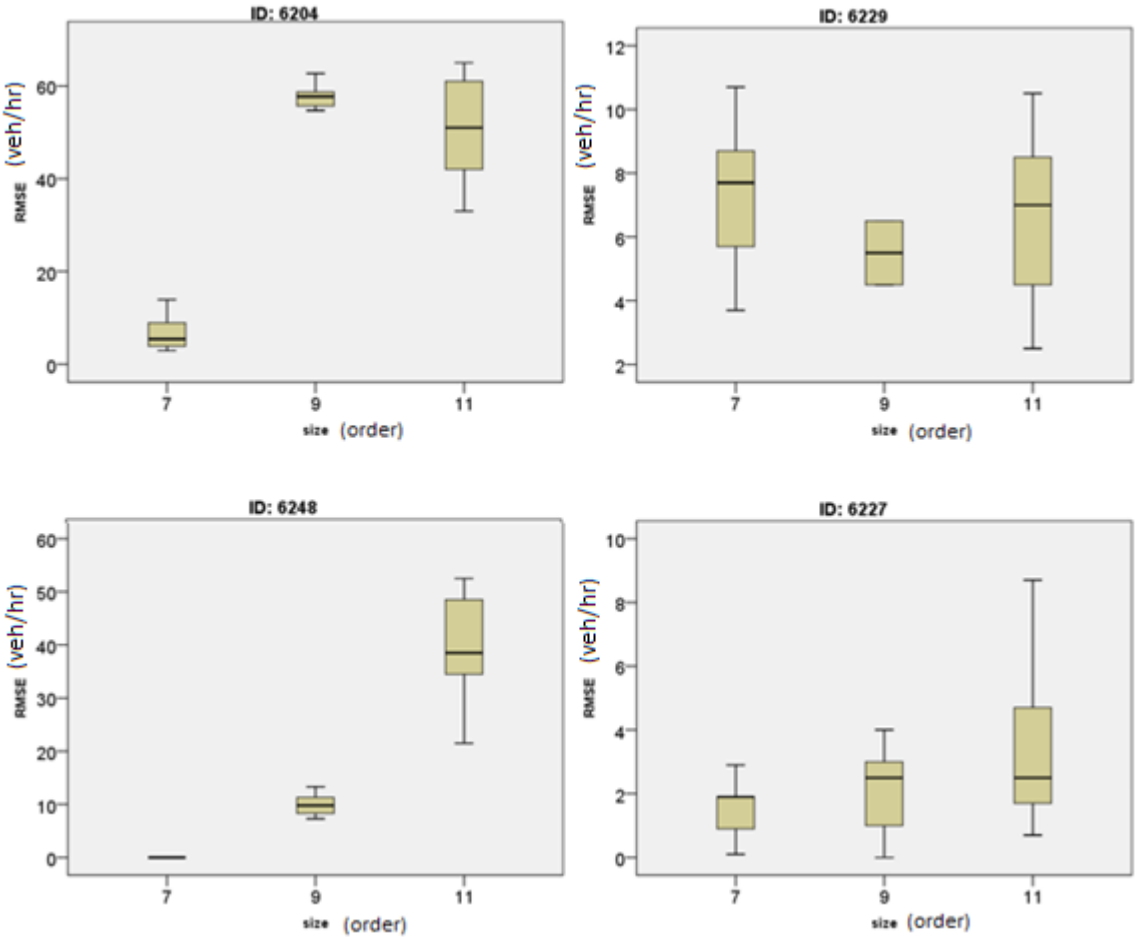


Figure 42 Boxplots for four selected links in the West MLK Network with a (2, 50%) configuration with varying sub-area size. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

Networks with (3, 75%) Configuration

Figure 43 through Figure 45 boxplot the RMSE values with increasing sub-area size across the three networks with three work zones and 75% capacity reduction. Unlike the inconsistent patterns seen in (1, 25%) and (2, 50%) (Figures 22-27), the RMSE values decrease when sub-area size increases (Figure 43 and Figure 44). However, we cannot find a significant and consistent pattern in Figure 45.

Figure 43and Figure 44are from the grid networks (Guadalupe Street and 7th Street), and both have similar work zones v/c ratios. Therefore, it seems reasonable to observe similar patterns from both networks and conclude that the sub-area size can impact target link when the work zones have significant capacity reduction in a grid network.

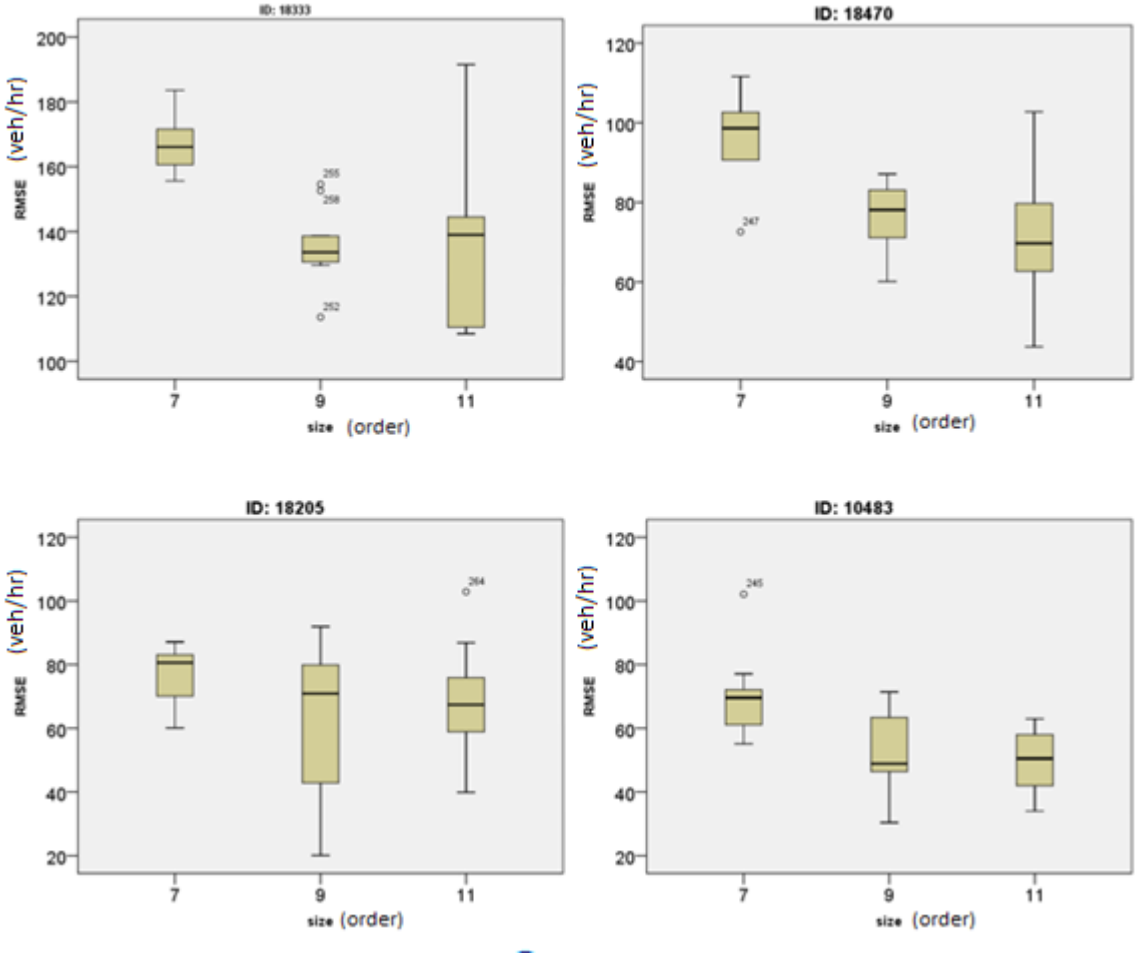


Figure 43 Boxplots for four selected links in the Guadalupe Street Network with a (3, 75%) configuration with varying sub-area size. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

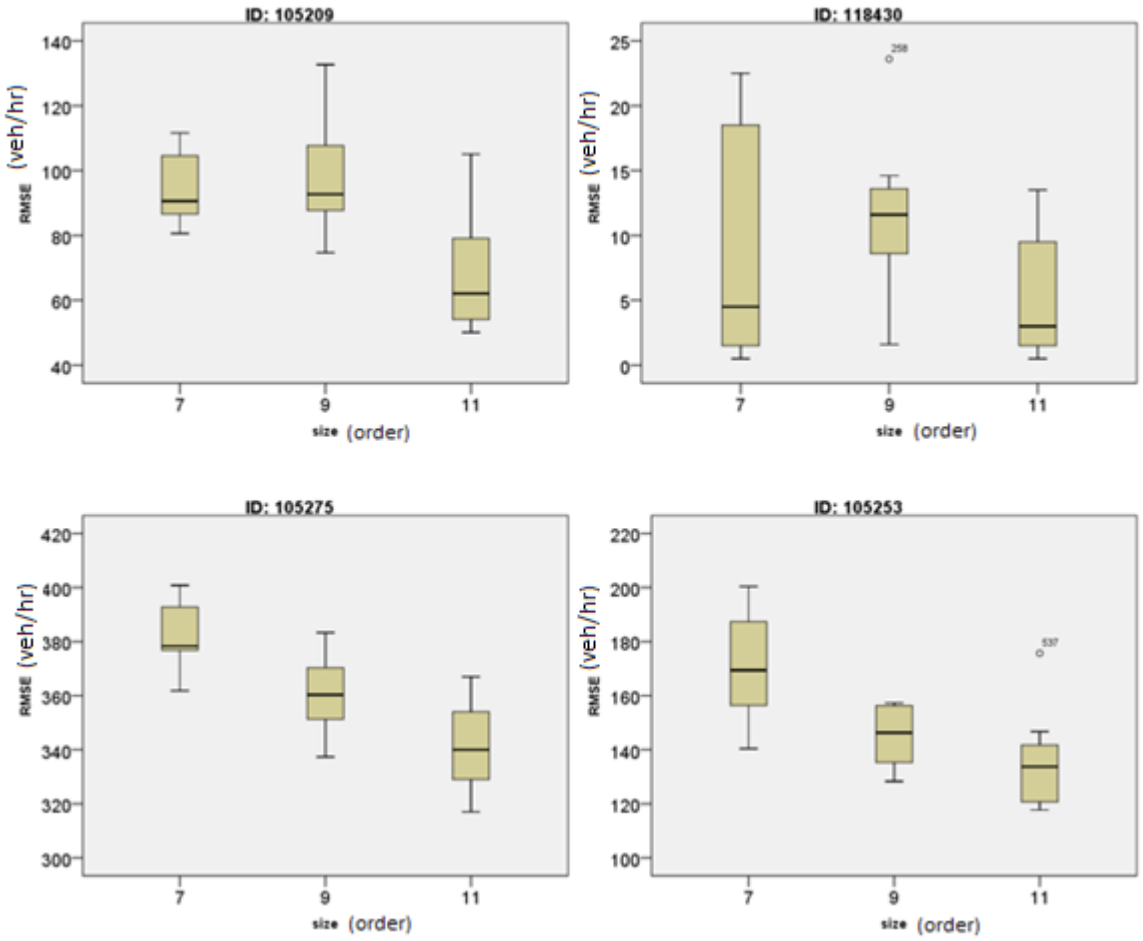


Figure 44 Boxplots for four selected links in the 7th Street Network with a (3, 75%) configuration with varying sub-area size. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

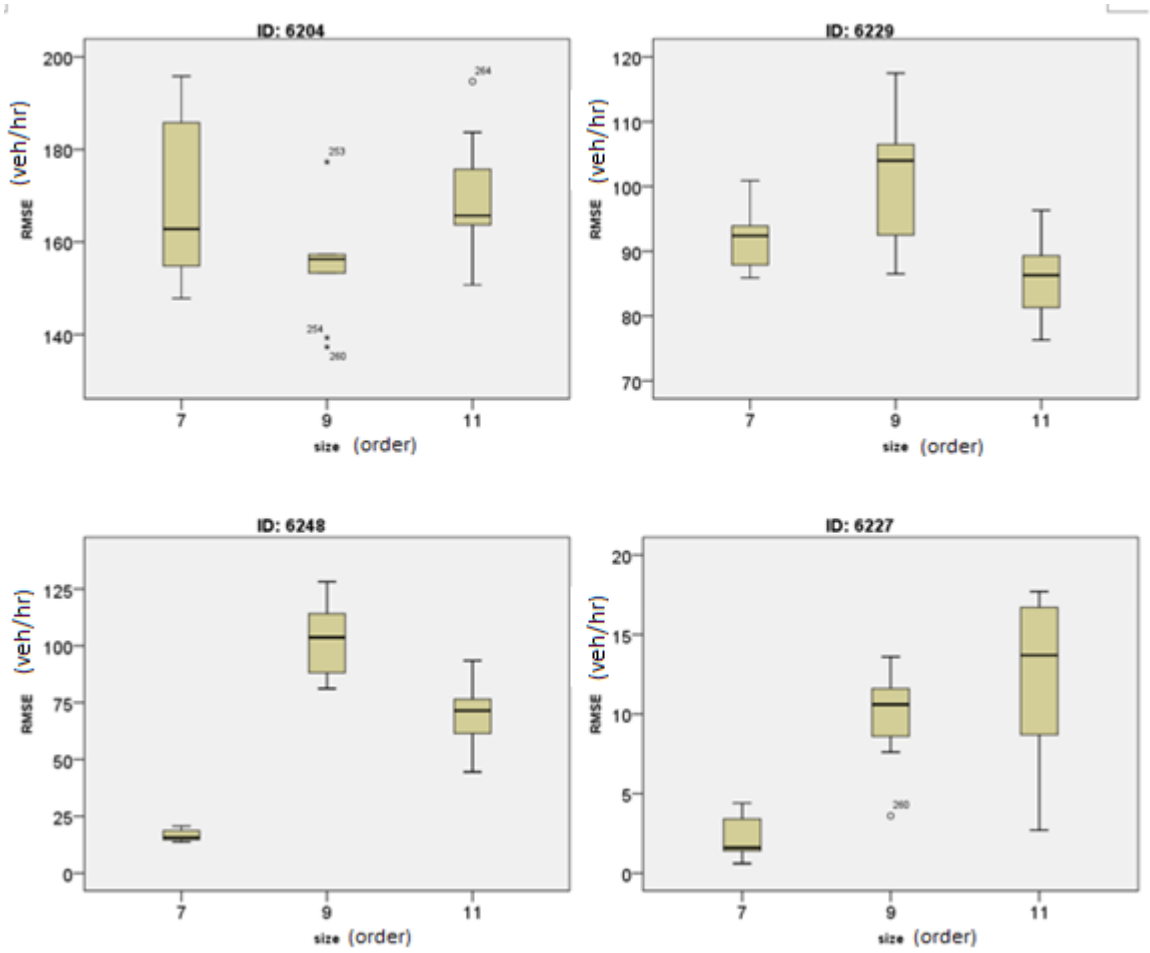


Figure 45 Boxplots for four selected links in the West MLK Network with a (3, 75%) configuration with varying sub-area size. The upper set are close to the work zones, and the lower set are far way from the work zones. The right pair are upstream of the work zones, and the left pair are downstream of the work zone.

Sub-Area Size Summary

This study boxplots the impact RMSEs when the network size increases across (1, 25%), (2, 50%), and (3, 75%) at Guadalupe Street, 7th Street, and West MLK.

With the (1, 25%) configuration, there was no consistent RMSE pattern observed in the Guadalupe Street network nor in the 7th Street network. However, in the West MLK network two links located far away with respect to the work zone showed an increase in RMSE with increasing sub-area size. With the (2, 50%) configuration, Guadalupe and West MLK did not show a consistent pattern across the target links, and two links close to the work zone in the 7th Street network showed an increase with increasing sub-area size. With the (3, 75%) configuration, three links in the Guadalupe and 7th Street networks showed a decrease in RMSE values when the sub-area increased. However, West MLK do not show the same pattern.

This study did not find a clear relationship between the sub-area size and RMSEs. Therefore, this variable was not included in the impacted regression model.

Summary

This study investigated the variables influencing RMSE values through a boxplot analysis. The investigated variables included the work zone capacity reduction, the work zone size, and the sub-area size. There are several findings, including:

1. **The capacity reduction in the work zone affects the target links.** When the capacity reduction increases, RMSE increases.
2. **The distance between the work zone and the target link influences RMSE values.** The RMSEs increases more when the work zone is close to the target link.
3. **The number of the work zones does not necessarily affect RMSE values.** The reducing the capacity in the first work zone causes the vehicles to reroute. However, the vehicles did not reroute when adding an additional work zone with the same capacity reduction.
4. **There is no consistent trend when increasing the sub-area size.** The Guadalupe and 7th Street Networks have similar trends when the capacity reduction is 75% and the work zone size is three. However, this study cannot

find a consistent trend when the capacity reduction and work zone sizes are less than three.

Due to these findings, this study considered the work zone's capacity reduction and the distance between the work zone and the target link as part of the impacted regression model. Conversely, this study will not consider work zone size and sub-area sizes as part of the impacted regression model.

5.4 Models

Adding a work zone causes a change in the network that can lead to vehicle rerouting, queue formation, and/or a change in travel times. To capture these effects, there are two types of networks, a DTA network and a microscopic traffic simulated network, that can be used. The DTA network captures vehicle rerouting as a result of having a work zone, and the microscopic traffic network captures detailed traffic information such as queue lengths or individual vehicle travel times. If the practitioner uses the information from a DTA simulation in a microscopic simulation, then the detailed information after vehicles reroute can be obtained.

However, an identical DTA network and the microscopic simulated network might not provide identical results. In a microscopic traffic simulator, driving behaviors are affected by the actions of the other vehicles. For example, a left-turning vehicle will stay in the left-side lane under the default setting. But if the left-turn traffic demand is more than the lane capacity, then the queue will spill back up to the upstream links. Conversely, vehicles in a DTA network stay in the links rather than in the lanes, so if the traffic demand, including left-turn, through, and right-turn, is over the capacity, the queue will not spill back to the upstream link. In this situation, the travel time in the microscopic traffic simulator might be much longer than in the DTA simulator. Thus, practitioners will have to calibrate the microscopic simulation network to get the same results from the DTA network. It is less time consuming and less challenging to calibrate a smaller network, however, the issue lies in determining an appropriate network size.

Traditionally, a hypothesis test has been used to define whether a target link is impacted or not. The hypothesis test process (Lehmann and Romano, 2005) is to first, gather the data from the base and the impacted scenarios; second, state the null and alternative

hypothesis and determine the appropriate statistical test (ex. T-test); third, select a significance level; fourth, compute the $t_{observe}$ from observations (for example, using t-test, the $t_{observe} = \frac{\bar{x}-u}{s/\sqrt{n}}$, where $\bar{x}$ is the sample mean, u is the population mean, s is standard deviation, and n is sample points); finally, determine if $t_{observe}$ rejects the null hypothesis. To follow this process requires both the base and impacted scenario observations, and becomes more time consuming as practitioners study more impacted scenarios.

This study developed two linear regression models for determining the impacted size more efficiently, one is called the impact model for estimating an impact RMSE, and the other is called the confidence interval model for estimating a base RMSE confidence interval. When the impact RMSE falls outside the base confidence interval, it is then reasonable to assume that the target link has experienced a significant impact from the work zone(s). Thus, this study can test whether, for example, a target link has a work zone by using the base RMSE value boundary and the RMSE from the impacted scenario.

The variables tested are the following,

- Flow: The link traffic flow from an individual simulation (unit: veh/hr),
- Order: The connected order of target links from the impacted link (Chen et al., 2012)
- Size: a variable indicating the size of the sub-area (Figure 46 shows the difference between Order and Size. The blue link is two links away from the work zone, therefore the blue link is a second order link relative to the target link. The gray links represent size of the sub-area = 3 because there are three links connected to the red work zone.),
- CapacityReduction: The percentage of capacity reduction on the impacted link represented as a portion of the original capacity level (unit: ratio),
- BaseZoneFlow: The sum of the capacities of each impacted link(s) divided by the number of links (unit: veh/hr),
- ImpactZoneFlow: BaseZoneFlow divided by (1- CapacityReduction) (unit: veh/hr). For example, if the average work zone's traffic flow is 300 in the base scenario and the capacity reduction is 25%. ImpactZoneFlow = 300/(1-0.25)= 400,
- BaseZoneVOverC: computed averages of all the work zones' volume to capacity ratio in the base scenario(unit: ratio),
- ImpactZoneVOverC: volume to capacity ratio calculated using the computed averages, divided by (1- CapacityReduction) (unit: ratio),
- MaxBaseZoneVoverC: The maximum v/c ratio for the work zone in the base scenario (unit: ratio),
- MaxBaseZoneFlow: The maximum traffic flow for the work zone in the base scenario (unit: veh/hr),
- ZoneCapacity: The work zones' total capacity (number lanes * lane capacity) divided by the number of impacted links (unit: veh/hr), and Figure 46 shows the red link capacity, and
- TotalCapacity: The target link's total capacity (lanes* lane capacity) divided by the number of impacted links (unit: veh/hr) (Figure 46 shows the blue link capacity).

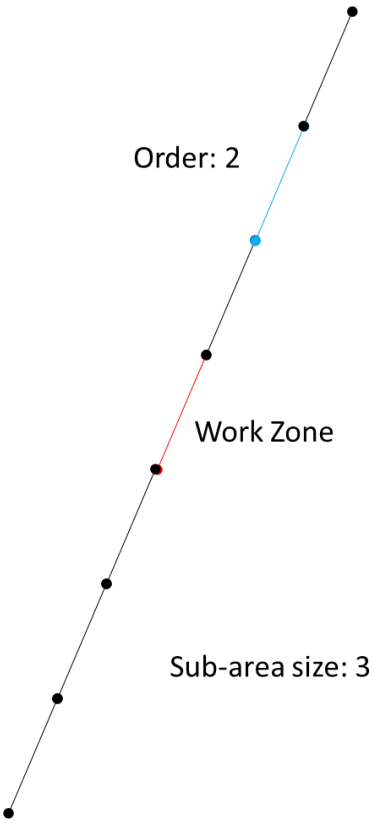


Figure 46 Example of a sub-area and a target link order

5.4.1 IMPACT RMSE MODEL

According to the boxplot analysis (Chapter 5.3), this study found there is a positive relationship between capacity reduction and impact RMSE values. Also, when traffic flow or v/c is large in the base scenario, the growing rate of the impacted scenario RMSE values is faster, so the relationship may not be linear. However, the impacted RMSE values and the confidence interval was used to determine whether the target link experiences a significant impact or not from the work zone. Here, the exact value is not necessary when the impact RMSE value is very large and falls outside of the confidence interval. Thus, this study has developed a linear regression model using capacity reduction as the independent variable for determining the impact RMSE values.

In addition, it was found that the impact RMSE values increase when the target link is close to the work zone (when capacity reduction is fixed). Thus, this evaluated adding the distance from the work zone to the regression model. Distance can be defined by feet or by order. This study followed Bringardener (2015) and used order as the unit to facilitate the after comparison. The preliminary model considers capacity reduction and the order (distance), which is shown in Table 11A.

Table 11 Model Specifications Tested Using Link Variables

	A		**B**		**C**		**D**	
Variable Tested	B	t	B	t	B	t	B	t
(Constant)	9	6.1	-13.863	-5.4369	140.808	100.020	5.252	3.293
Order	-11.29	-66.5	-11.285	-66.705	-11.286	-63.060	-11.755	-63.232
CapacityReduction	1.583	66.17	1.58265	66.4304	x	x	1.583	66.253
BaseZoneFlow	x	x	0.02479	10.9758	x	x	x	x
ImpactZoneFlow	x	x	x	x	-0.114	-49.127	x	x
BaseZoneVOverC	x	x	x	x	x	x	23.105	6.170
ImpactZoneVOverC	x	x	x	x	x	x	x	x
MaxBaseZoneVoverC	x	x	x	x	x	x	x	x
MaxBaseZoneFlow	x	x	x	x	x	x	x	x
ZoneCapacity	x	x	x	x	x	x	x	x
BaseFlow	x	x	x	x	x	x	x	x
Adjusted R square	0.364		0.369		0.294		0.366	

	E		**F**		**G**		**H**	
Variable Tested	B	t	B	t	B	t	B	t
(Constant)	7.247	3.009	-120.773	-22.658	-357.127	-23.233	-183.542	-32.238
Order	-11.761	-67.468	-10.999	-65.946	-10.997	-65.768	-10.418	-64.177
CapacityReduction	x	x	1.583	67.536	1.583	67.440	1.583	69.878
BaseZoneFlow	x	x	x	x	x	x	x	x
ImpactZoneFlow	x	x	x	x	x	x	x	x
BaseZoneVOverC	x	x	x	x	x	x	x	x
ImpactZoneVOverC	102.848	58.620	x	x	x	x	x	x
MaxBaseZoneVoverC	x	x	x	x	696.302	24.299	x	x
MaxBaseZoneFlow	x	x	0.120	25.296	x	x	0.178	36.299
ZoneCapacity	0.007	18.595	x	x	0.040	23.410	-0.007	-19.726
BaseFlow	x	x	x	x	x	x	0.040	32.704
Adjusted R square	0.332		0.389		0.388		0.430	

The work zone's average traffic flow or v/c in the base scenario also influences the RMSE values resulting from the impacted scenario. The impact RMSE value increases more when the traffic flow or v/c ratios are higher. Thus, this study evaluated adding the work zones' average traffic flow and v/c ratios to the model (Table 11B and 11D, respectively). Adding in the work zone's average traffic flow showed a better fit than adding in the work zone's v/c ratio to this database, according to higher adjusted R square.

The work zone's both traffic flow or v/c in the base scenario and the capacity reduction was found to impact the RMSE values from the target links. But, the work zone's

average traffic flow or v/c and capacity reduction does not represent the impacted traffic condition very well. The adjusted R square increases 0.05 or 0.02 for adding the work zone's average traffic flow or v/c. It might provide a better fit to consider the work zone's traffic flow and capacity reduction, so this study replaces the traffic flow or v/c ratio and capacity reduction with the impacted traffic flow and impacted v/c ratio (Table 11C and 9E). However, Table 11B and D still show better results.

The results from (B) through (E) led this study to focus on the work zone's condition in the base scenario because using the impacted work zone traffic flow or v/c did not work well. The target link will be influenced by the work zone, which is most significantly impacted by the capacity reduction. Thus, using the max work zone flow rather than the average work zone flow works better (Table 11F). In addition, this study replaced the average v/c (Table 11D) with the max work zone v/c and the zone capacity (Table 11G). The adjusted R squares are similar between (F) and (G), but the number of variables in (F) is less than in (G). Therefore, comparing (A) through (F) shows that (F) was best option.

Finally, this study added the link base traffic flow to (F), and the result is shown in (H), which provided the best adjusted R square, 0.430. Thus, this study uses (H) as the impact model. The summary of the entire Final Impact Model for determining the impacted RMSE values is presented in Table 12.

Table 12 Results from the Final Impact Model

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
0.656	0.430	0.430	57.297

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	38008240.948	5	7601648.190	2315.508	0.000
Residual	50406095.109	15354	3282.929		
Total	*88414336.057*	*15359*			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-183.542	5.693		-32.238	0.000
Order	-10.418	0.162	-0.395	-64.177	0.000
CapacityReduction	1.583	0.023	0.426	69.878	0.000
MaxBaseZoneFlow	0.178	0.005	0.238	36.299	0.000
BaseFlow	0.040	0.001	0.241	32.704	0.000
ZoneCapacity	-0.007	0.000	-0.139	-19.726	0.000

The equation for the impact model is

$$Impact_{RMSE} = -183.542 - 10.418 * Order + 1.583 * CapcityReduction + 0.178 * MaxBaseZoneFlow - 0.007 * ZoneCapacity + 0.04 * BaseFlow \quad (23).$$

5.4.2 CONFIDENCE INTERVAL MODEL

This study also built a linear regression model for estimating the target link's RMSE confidence interval. If the estimated impact RMSE from the impact model is outside of the estimated RMSE confidence interval, then one can assume that the target link is impacted by a work zone.

Similar variables were added in the confidence interval models such as, the target link flow, the max work zone's flow, and the target link order, which led to an adjusted R

square = 0.294. Following intuition, adding the sub-area size and the target link total capacity in this model, increased the adjusted R square to 0.459. The model is,

Model Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate
.678[a]	0.459	0.459	2.215814474580340

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Regression	21311.782	5	4262.356	868.126	0
Residual	25108.890	5114	4.910		
Total	46420.672	5119			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-6.958	0.436		-15.968	0.000
flow	0.001	0.000	0.207	16.478	0.000
Order	-0.284	0.012	-0.272	-24.562	0.000
MaxZoneFlow	0.003	0.000	0.099	9.125	0.000
TotalCapacity	0.001	0.000	0.416	32.762	0.000
Size	0.426	0.020	0.228	21.628	0.000

$$Confidence\ Interval\ = -6.958 - 0.284 * Order + 0.426 * Sub - area\ Size + 0.01 * flow + 0.003 * MaxZoneFlow + 0.001 * TotalCapacity \quad (23).$$

5.4.3 RESULT AND ANALYSIS

This study developed an impact RMSE model and a confidence interval model for defining whether there is an impact from a work zone on a target link. To verify the models' ability, Table 13 through 15 present the comparison between the hypothesis test and the model results ranging from 25 to 75 percent capacity reduction across sub-area size 7 through 11, and one to three work zones. In these tables, the hypothesis test used was the T-test to compare whether the impact RMSE mean and base RMSE mean

were identical. If the both means were identical, then the cell reads "No"; otherwise, the cell reads "Impact". The column titled *Model Result* shows the result from impact and the confidence interval models. If the impacted value is larger than the confidence interval value, then the cell shows "impact"; otherwise, it shows "No". The column titled *Compare* shows the result between the Hypothesis Test and the Model Result. If the both results are the same, the cell shows "O"; otherwise, it shows "X".

Overall, the models work well in these locations, especially for the West MLK network. Guadalupe does not present good results when the capacity reduction is 50%. The mild case is represented by 25% reduction and the extreme case is 75% reduction, but it is difficult to define when the v/c or traffic flow is small. For the 7th Street network, the results are not good when the network size is large. This is likely because 7th Street is close to the freeway, therefore when the network size becomes larger, part of freeway system is included. Including the freeway adds more complexity because most vehicles prefer to use the on-ramp for the shortest path even when there is a work zone, so the vehicles do not reroute as expected.

Table 13 Comparison of the Final Impact Model and Confidence Interval Model for networks with a (7, 1) configuration

Guadalupe Street Network (7,1)

	Hypothesis Test[1]			Model Result			Comparison		
linkid	**25**	**50**	**75**	**25**	**50**	**75**	**25**	**50**	**75**
18428	No	No	Impact	Impact	Impact	Impact	X	X	O
5225	No	Impact	Impact	No	Impact	Impact	**O**	**O**	**O**
5146	No	Impact	No	No	Impact	Impact	O	O	X
18333	No	No	Impact	No	Impact	Impact	O	X	O
5189	No	Impact	No	No	Impact	Impact	O	O	X
18474	No	No	Impact	No	Impact	Impact	O	X	O
18470	No	Impact	Impact	No	Impact	Impact	**O**	**O**	**O**
5219	No	No	Impact	No	Impact	Impact	O	X	O
4608	No	No	Impact	No	Impact	Impact	O	X	O
18205	No	No	Impact	No	Impact	Impact	O	X	O
114984	No	No	Impact	No	Impact	Impact	O	X	O
104883	No	No	Impact	No	Impact	Impact	O	X	O
18381	No	Impact	Impact	No	Impact	Impact	**O**	**O**	**O**
18500	No	No	No	No	Impact	Impact	O	X	X

Table 13 Continued. Comparison of the Final Impact Model and Confidence Interval Model for networks with a (7, 1) configuration

7th Street Network (7,1)

	Hypothesis Test			Model Result			Comparison		
linkid	**25**	**50**	**75**	**25**	**50**	**75**	**25**	**50**	**75**
118172	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105201	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105142	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
118440	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
118450	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105209	No	Impact	Impact	Impact	Impact	Impact	X	O	O
105275	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105162	No	No	Impact	Impact	Impact	Impact	X	X	O
116041	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105147	No	No	Impact	No	Impact	Impact	O	X	O
105268	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105253	No	Impact	Impact	Impact	Impact	Impact	X	O	O
118430	No	No	Impact	No	Impact	Impact	O	X	O
114720	No	No	Impact	No	Impact	Impact	O	X	O
117617	No	No	No	No	Impact	Impact	O	X	X

West MLK Network (7, 1)

	Hypothesis Test			Model Result			Comparison		
linkid	**25**	**50**	**75**	**25**	**50**	**75**	**25**	**50**	**75**
6214	No	Impact	Impact	Impact	Impact	Impact	X	O	O
6232	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
6212	No	Impact	Impact	Impact	Impact	Impact	X	O	O
6204	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18514	No	No	Impact	Impact	Impact	Impact	X	X	O
18508	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18265	No	No	Impact	Impact	Impact	Impact	X	X	O
14711	No	Impact	Impact	Impact	Impact	Impact	X	O	O
6229	No	No	Impact	No	Impact	Impact	O	X	O
6252	No	Impact	Impact	Impact	Impact	Impact	X	O	O
6248	No	No	Impact	Impact	Impact	Impact	X	X	O
18512	No	No	No	No	Impact	Impact	O	X	X
18275	No	Impact	Impact	No	Impact	Impact	O	O	O
18561	No	No	Impact	No	Impact	Impact	O	X	O
6227	Impact	No	No	No	Impact	Impact	X	X	X
5885	No	Impact	Impact	No	Impact	Impact	O	O	O
18283	No	Impact	Impact	No	Impact	Impact	O	O	O

1. Y = Accept Ho: $\mu 1 = \mu 2$; Impact = Reject Ho conclude Ha: $\mu 1 \neq \mu 2$

Table 14 Comparison of the Final Impact Model and Confidence Interval Model for networks with a (9, 2) configuration

Guadalupe Street Network (9, 2)

	Hypothesis Test			Model Result			Comparison		
linkid	**25**	**50**	**75**	**25**	**50**	**75**	**25**	**50**	**75**
18428	No	Impact	Impact	Impact	Impact	Impact	X	O	O
5146	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
5225	No	No	Impact	No	Impact	Impact	O	X	O
5189	Impact	Impact	Impact	No	Impact	Impact	X	O	O
18333	No	No	Impact	No	Impact	Impact	O	X	O
18470	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18474	No	No	Impact	No	Impact	Impact	O	X	O
4608	No	Impact	Impact	Impact	Impact	Impact	X	O	O
5219	No	No	Impact	No	Impact	Impact	O	X	O
114984	No	No	Impact	No	Impact	Impact	O	X	O
18205	No	No	Impact	No	Impact	Impact	O	X	O
104883	No	No	Impact	No	Impact	Impact	O	X	O
18500	Impact	No	Impact	No	Impact	Impact	X	X	O
18381	No	No	Impact	No	Impact	Impact	O	X	O
18200	Impact	No	Impact	No	Impact	Impact	X	X	O
5229	Impact	No	Impact	No	No	Impact	X	O	O
4258	No	No	No	No	No	Impact	O	O	X
4787	No	No	Impact	No	Impact	Impact	O	X	O
117612	No	No	No	No	Impact	Impact	O	X	X

7th Street Network (9, 2)

	Hypothesis Test			Model Result			Compare		
linkid	**25**	**50**	**75**	**25**	**50**	**75**	**25**	**50**	**75**
118172	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105142	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105201	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
118440	No	Impact	Impact	Impact	Impact	Impact	X	O	O
118450	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105209	No	Impact	Impact	Impact	Impact	Impact	X	O	O
105275	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105162	Impact	Impact	Impact	No	Impact	Impact	X	O	O
116041	No	No	No	Impact	Impact	Impact	X	X	X
105147	No	Impact	Impact	No	Impact	Impact	O	O	O
105268	No	Impact	No	Impact	Impact	Impact	X	O	X
105253	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
118430	No	No	Impact	No	Impact	Impact	O	X	O
114720	No	No	No	No	Impact	Impact	O	X	X
117617	No	No	No	No	No	Impact	O	O	X
104624	No	No	No	No	No	Impact	O	O	X
104906	No	No	No	No	No	Impact	O	O	X

Table 14 Continued. Comparison of the Final Impact Model and Confidence Interval Model for networks with a (9, 2) configuration

West MLK Network (9, 2)

	Hypothesis Test			Model Result			Compare		
linkid	**25**	**50**	**75**	**25**	**50**	**75**	**25**	**50**	**75**
6214	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
6212	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
18514	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
6232	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18265	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
6204	No	Impact	Impact	Impact	Impact	Impact	X	O	O
6229	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18508	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18275	No	Impact	Impact	Impact	Impact	Impact	X	O	O
14711	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18561	No	Impact	Impact	No	Impact	Impact	O	O	O
6252	No	Impact	Impact	Impact	Impact	Impact	X	O	O
6227	Impact	No	Impact	No	Impact	Impact	X	X	O
5885	Impact	Impact	Impact	No	Impact	Impact	X	O	O
6248	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18283	No	Impact	Impact	No	Impact	Impact	O	O	O
18512	No	Impact	Impact	No	Impact	Impact	O	O	O
14704	No	Impact	Impact	No	Impact	Impact	O	O	O
6255	No	No	No	No	Impact	Impact	O	X	X
6012	Impact	Impact	Impact	No	No	Impact	X	X	O
6009	No	No	Impact	No	No	Impact	O	O	O

Table 15 Comparison of the Final Impact Model and Confidence Interval Model for networks with a (11, 3) configuration

Guadalupe Street Network (11,3)

	Hypothesis Test			Model Result			Compare		
linkid	**25**	**50**	**75**	**25**	**50**	**75**	**25**	**50**	**75**
18428	No	Impact	Impact	Impact	Impact	Impact	X	O	O
5146	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
5225	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
5189	Impact	Impact	Impact	No	Impact	Impact	X	O	O
18333	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18470	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18474	No	No	Impact	No	Impact	Impact	O	X	O
4608	Impact	Impact	Impact	No	Impact	Impact	X	O	O
5219	No	No	Impact	No	Impact	Impact	O	X	O
114984	No	No	Impact	No	Impact	Impact	O	X	O
18205	No	No	Impact	No	Impact	Impact	O	X	O
104883	No	No	Impact	No	Impact	Impact	O	X	O
18500	No	No	Impact	No	Impact	Impact	O	X	O
18381	No	No	Impact	No	Impact	Impact	O	X	O
18200	No	No	Impact	No	Impact	Impact	O	X	O
5229	No	No	No	No	Impact	Impact	O	X	X
4258	No	No	No	No	No	Impact	O	O	X
4787	No	No	No	No	Impact	Impact	O	X	X
117612	No	No	No	No	No	No	O	O	O

Table 15 Continued. Comparison of the Final Impact Model and Confidence Interval Model for networks with a (11, 3) configuration

7th Street Network (11,3)

	Hypothesis Test			Model Result			Compare		
linkid	**25**	**50**	**75**	**25**	**50**	**75**	**25**	**50**	**75**
118172	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105201	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105142	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
118440	No	Impact	Impact	Impact	Impact	Impact	X	O	O
118450	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
105209	No	Impact	Impact	Impact	Impact	Impact	X	O	O
105275	No	Impact	Impact	Impact	Impact	Impact	X	O	O
105162	No	Impact	Impact	Impact	Impact	Impact	X	O	O
116041	No	Impact	Impact	Impact	Impact	Impact	X	O	O
105147	No	Impact	Impact	No	Impact	Impact	O	O	O
105268	No	Impact	Impact	Impact	Impact	Impact	X	O	O
105253	No	Impact	Impact	Impact	Impact	Impact	X	O	O
118430	Impact	No	No	No	Impact	Impact	X	X	X
114720	No	No	No	Impact	Impact	Impact	X	X	X
117617	No	No	No	No	Impact	Impact	O	X	X
104624	No	No	No	No	No	Impact	O	O	X
104906	No	No	No	No	No	Impact	O	O	X
118315	No	No	No	No	No	Impact	O	O	X
118348	No	No	No	No	No	Impact	O	O	X

Table 15 Continued. Comparison of the Final Impact Model and Confidence Interval Model for networks with a (11, 3) configuration

West MLK Street Network (11,3)

	Hypothesis Test			Model Result			Compare		
linkid	**25**	**50**	**75**	**25**	**50**	**75**	**25**	**50**	**75**
6214	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
6212	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18514	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
6232	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
18265	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
6204	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
6229	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18508	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
18275	No	Impact	Impact	No	Impact	Impact	O	O	O
14711	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
18561	No	No	Impact	No	Impact	Impact	O	X	O
6252	Impact	Impact	Impact	Impact	Impact	Impact	O	O	O
6227	Impact	Impact	Impact	No	Impact	Impact	X	O	O
5885	No	No	Impact	No	Impact	Impact	O	X	O
6248	No	Impact	Impact	Impact	Impact	Impact	X	O	O
18283	No	Impact	Impact	No	Impact	Impact	O	O	O
18512	No	No	Impact	No	Impact	Impact	O	X	O
14704	No	Impact	Impact	No	Impact	Impact	O	O	O
6255	No	No	Impact	No	Impact	Impact	O	X	O
6188	No	No	Impact	No	Impact	Impact	O	X	O
6338	No	No	Impact	No	No	Impact	O	O	O
6334	No	No	No	No	No	Impact	O	O	X
6012	No	No	Impact	No	No	Impact	O	O	O
6009	No	No	Impact	No	No	Impact	O	O	O

Table 16 displays the accuracy rate of the Final Impact Model and Confidence Interval Model for the network configurations. This table shows that the accuracy rate increases when the capacity reduction increases. When the capacity reduction is 25%, the accuracy rate is between 40% and 55%; as 50% capacity reduction, the accuracy rate is between 70% and 85%; and when 75% capacity reduction, the accuracy rate is between 90% and 100%. But the accuracy rate is stable without considering the capacity reduction. Most of the accuracy rate is between 70% and 80% across the sub-area size and work zone size. This table proves that the vehicle paths are significantly impacted by the capacity reduction rather than the sub-area size and the work zone sizes.

Table 16 Comparison of the Final Impact Model and Confidence Interval Model for network configurations, including capacity reduction, sub-area size, and work zone size.

		Capacity Reduction (%)		
		25	50	75
Sub-area size (order)	7	54.8%	83.8%	97.0%
	9	46.3%	78.8%	97.3%
	11	48.6%	73.0%	92.8%
Work zone size (num)	1	44.6%	75.7%	96.2%
	2	48.0%	77.0%	94.3%
	3	55.8%	81.3%	96.0%
		Sub-area size (order)		
		7	9	11
Work zone size (num)	1	76.9%	72.6%	68.3%
	2	74.9%	74.9%	70.1%
	3	83.3%	74.8%	75.9%

CHAPTER 6. DTA SUBAREA DEVELOPER

DTA simulation can provide more accurate traffic volumes than STA, but DTA is computationally expensive. DTA simulation requires dealing with multiple OD matrices, since this type of traffic assignment accounts for the fact that traffic conditions change over time. However, using a subnetwork is a potential solution for reducing the computational cost.

Currently, the process for extracting a part of an urban network (a subnetwork) requires defining the size order and coding that selected portion of the whole network. The size order is defined as the number links connected to the modified impacted links that should be included in the subnetwork. Subnetwork links are identified by size order using ArcGIS, however, ArcGIS treats link intersections in only two dimensions, so link crossings of elevated freeways are erroneously identified as at-grade intersections. Hence, a subnetwork extracted using ArcGIS might contain non-existent at-grade intersection and might be larger than desired.

Furthermore, the current process is time consuming and human mistakes can be easily made. The entire network in ArcGIS contains four layers: a link layer, a node layer, a connector layer, and a centroid layer. The subnetwork links are first isolated from the link layer using the size order defined by the user. The subnetwork nodes representing intersections of the subnetwork links are then isolated from the node layer. The subnetwork connectors and the subnetwork centroids are also isolated in the same way. After that, the sub-link and node data are imported into the copied network in the VISTA database through SSH. Then, the user must modify the code for developing sub OD matrices according to the sub-link data, sub-node data, and the vehicle paths that evolved for the entire network. The subnetwork database is created by running the modified code.

The DTA subnetwork developer automates the process of extracting a subnetwork from a full network with an optimized subnetwork size model. The program contains two

subsystems, one for extracting the subnetwork geometry and the other for creating a subnetwork database through SSH and VISTA. The program needs 1) the link descriptions for the entire network, including link IDs, upstream and downstream node ID's, and 2) the link properties, including capacity, number of lanes, and speed limit. In order to simulate changes to the network using the automatic subnetwork tool, one must describe the impacted scenario including how the network was changed with closures or other modifications. The user must enter impacted link IDs, the capacity reduction of the impacted links, the acceptable RMSE, and the regional network from which data will remotely be extracted using SSH (Figure 47).

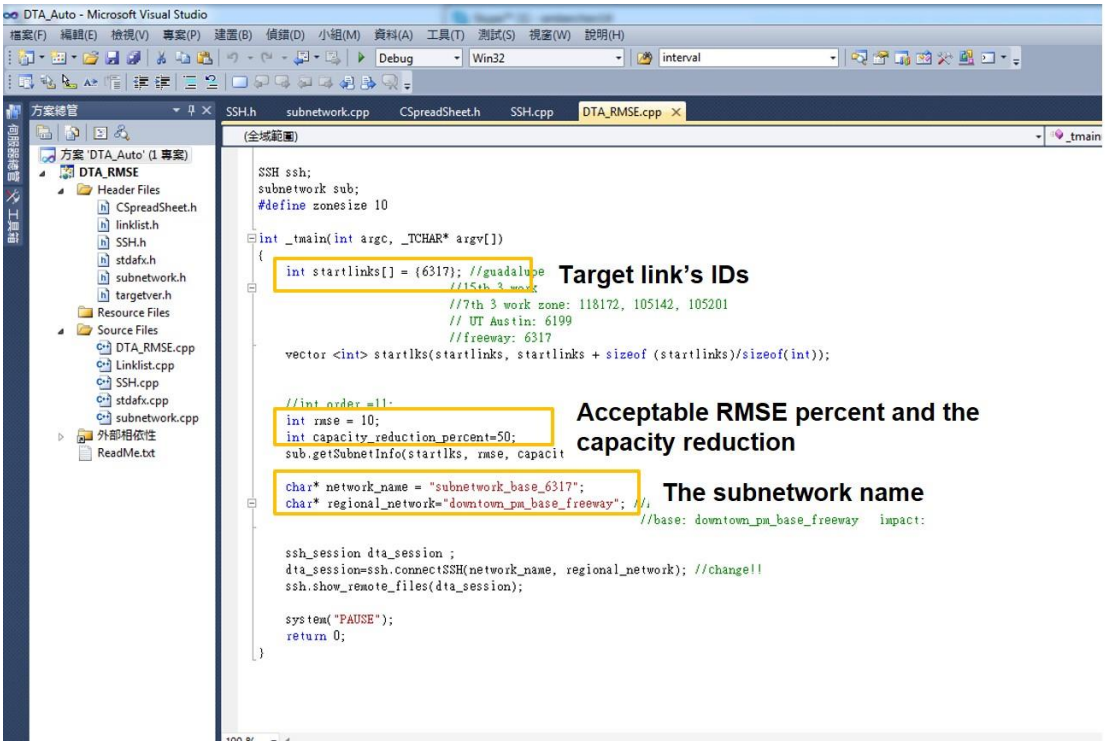


Figure 47 The modified subnetwork program inputs

Then, the program runs the procedure found in Figure 48. In this study, Section 6.1 will introduce how this program determines the subnetwork size and how to extract the subnetwork, Section 6.2 will introduce how the program develops sub Origin-Destination (OD) matrices and the subnetwork database in VISTA; Section 6.3 will present the downtown Austin network as a case study.

Automatic Program Procedure

Input: Regional network and impacted scenario data

First subsystem: The subnetwork's link and node data

Second subsystem:

1. SSH connection
2. The subnetwork's OD matrix

Output: The subnetwork database

Figure 48 Automatic Program Procedure

6.1 First subsystem

In the first step, this program refers to Bringardner's (2015) models to calculate the optimal subnetwork size with a user chosen acceptable error percentage. The RMSEs (Equation (14) and Equation (15) in Chapter 2) are only affected by the size order, since the user defines the other parameters (percent capacity reduced, number of impacted links, etc.). The size order in Equation (14) and Equation (15) starts from one and increases until the RMSEs in the base and the impacted model are equal. Once they are equal, the optimal subnetwork size is determined.

In the second step, the program searches for link IDs based on the determined subnetwork size. The connected links are identified by node ID numbers because connected links share the same node. This method helps avoid the previously encountered problem of incorrectly identifying grade separations as intersections, as seen in ArcGIS before. After isolating the subnetwork's link and node data, the connecters and the centroids attached to the subnetwork area also identified. Finally, the link data including links, nodes, centroids and connectors for the subnetwork is complete.

6.2 Second subsystem

In this subsystem, SSH copies the regional network database to the subnetwork database, including the full network OD matrix and the vehicle tracking data. Then, this program replaces the entire network's link and node data with the subnetwork data. The modified JavaScript code extracts and replaces the full network OD matrix with the sub OD matrix. The modification includes the location of the regional OD matrix and the

subnetwork link data. After building the subnetwork geometry and the sub OD matrix in VISTA, the development of the subnetwork database is finished. Figure 49 shows the output after using the automatic program; the included links are size order three and there is a 50 percent capacity reduction on several links along Guadalupe Street (orange line).

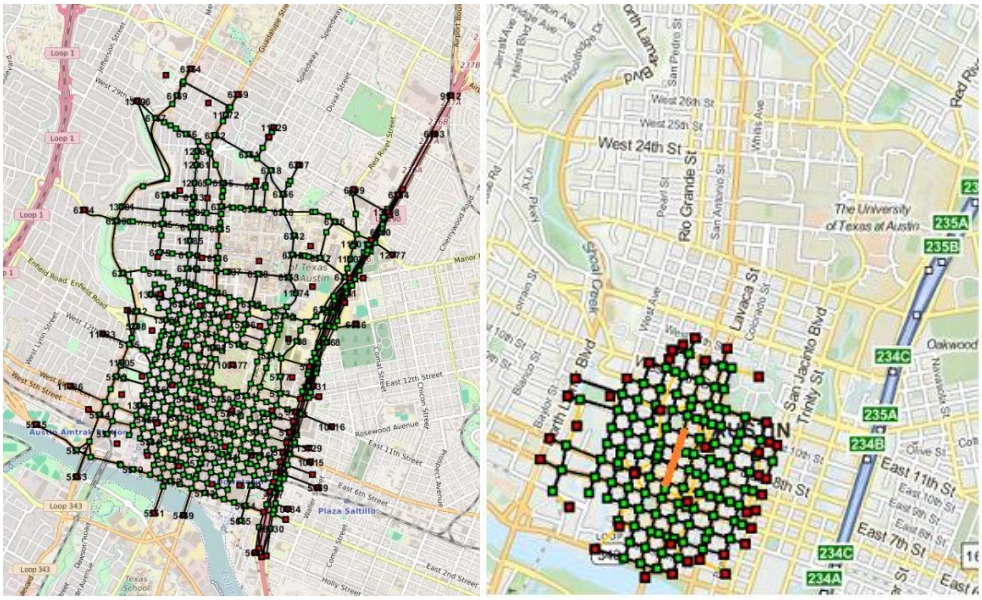


Figure 49 Example of a subnetwork developed with the automatic program. To the left, the entire network and to the right, the subnetwork.

6.3 Case Study

This study uses the Downtown Austin network as the case study. The work zones/impacted links (Guadalupe, 7th, and 15th) and the order sizes used (7, 9, 11) are identical to what was used in Bringardener's (2015) work to validate that the automatic program followed his procedure (Figure 50).

The link data is shown in Table 17. The numbers of links in the subnetwork created with Bringardener's (2015) work and the automatic program are identical for the Guadalupe Street scenario, but are different for the 7th Street and 15th Street scenarios. The subnetwork only includes urban streets in the Guadalupe Street scenario. The 7th Street and 15th Street scenarios included freeway links when the order size = 7. Since ArcGIS cannot see the elevation of the crossroads, the number of links included under order size 7 in ArcGIS is more than the amount included when using the automatic program. The difference between the number of links in Bringardener's work and the

new automatic program increases in the 7th and 15th Street scenarios as the work zone size and the size order increases.

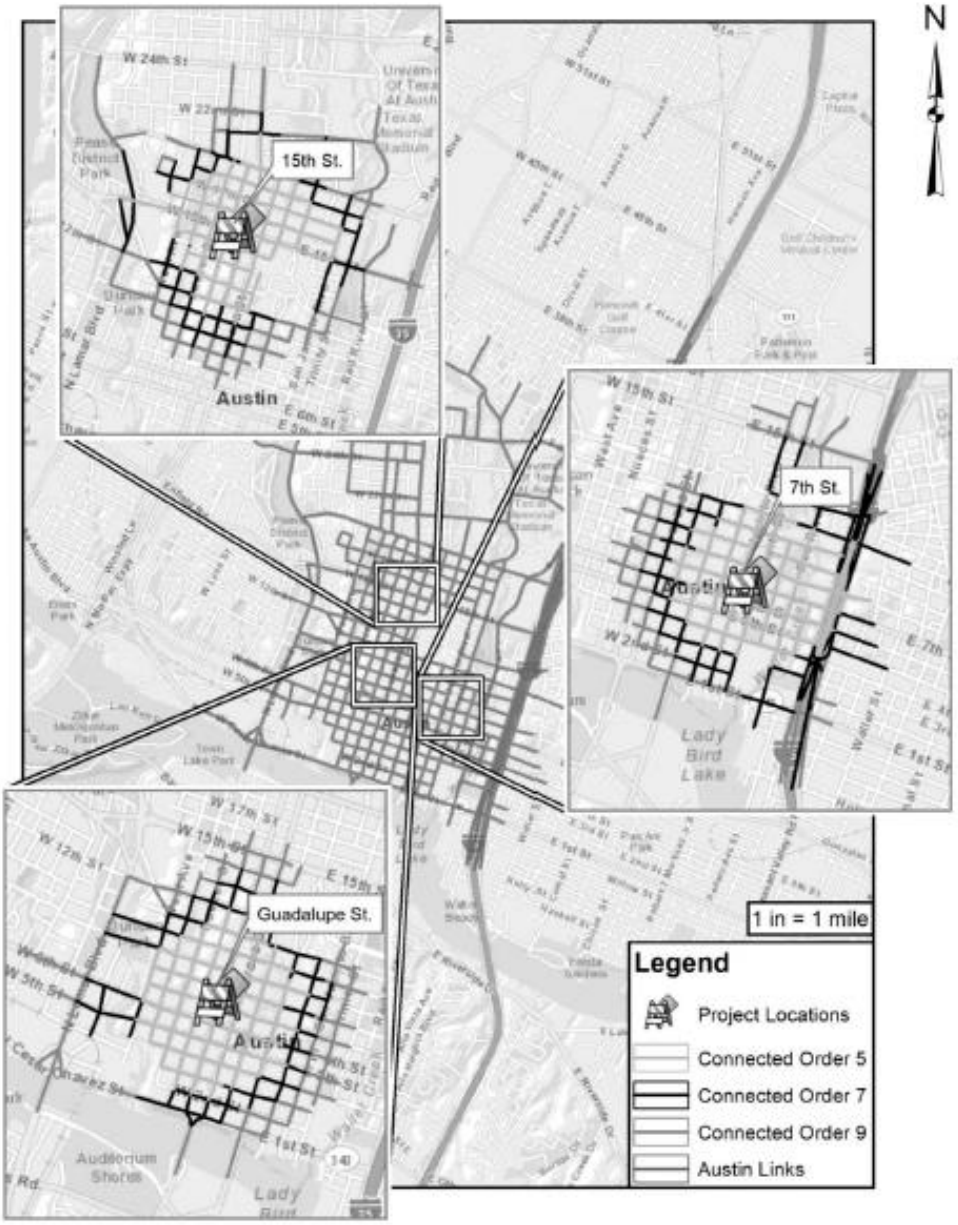


Figure 50 Case study locations (Bringardner, 2015)

Table 17 Number of links needed based on size order for 15th, 7th, and Guadalupe Street

Bringardner (2015)				Auto Program			Difference		
15th	**Order**			**Order**			**Order**		
work zone	7	9	11	7	9	11	7	9	11
1	396	595	--	388	588	--	8	7	--
2	425	645	848	420	638	825	5	7	23
3	442	660	876	437	649	847	5	11	29
7th	**Order**			**Order**			**Order**		
work zone	7	9	11	7	9	11	7	9	11
1	371	572	--	362	559	--	9	13	--
2	404	584	714	--	568	694	--	16	20
3	420	598	--	407	582	--	13	16	--
Guadalupe	**Order**			**Order**			**Order**		
work zone	7	9	11	7	9	11	7	9	11
1	368	539	--	368	539	--	0	0	--
2	402	569	807	402	569	807	0	0	0
3	448	625	824	448	625	824	0	0	0

6.4 Summary

This study implements a program that automatically develops a subnetwork database in VISTA. The program was created and validated with Bringardener's (2015) RMSE models. Currently, inputs are the full network and impacted scenario data, including the impacted link, its capacity reduction, and the acceptable error. The program output is the subnetwork database. After we develop the optimization models using the traditional optimization process, the models will replace Bringardener's (2015) RMSE models in this program.

This program has the ability to simplify the process and to reduce the possibility of manual mistakes. The problem of elevated crossroads being treated as at-grade intersections has been solved by removing the need to use ArcGIS for the subnetwork extraction. Additionally, the reduction in simulation time allows for more opportunities to discuss and analyze more alternatives to improve traffic conditions in a shorter about of time. Furthermore, new avenues of research on subnetwork analysis could be opened

up due to the convenience of using this program.

CHAPTER 7 CORSIM TRANSFORMER

A dynamic traffic assignment (DTA) simulator provides traffic demand and vehicle routings, but does not provide the quality of traffic flow on links composing the network region. DTA can provide a good picture of the demand resulting from modification to the network. On the other hand, a microscopic traffic simulator, such as CORSIM, can provide information about the quality of traffic flow on links, but cannot provide information about the change in traffic demand. While CORSIM researchers and users can see the immediate impact to traffic conditions (such as speed, travel time, and link flow) caused by adjusting the network geometry, they cannot see the resulting rerouting. Therefore, DTA is appropriate for analyzing long-term changes in network geometry but a microscopic simulator, is more appropriate for analyzing short-term network changes.

It would be ideal for researchers or transportation engineers to have an optimal solution that works for both short-term and long-term situations, but many only have a DTA network coded. There are situations where it would be useful to develop an identical network in a microscopic traffic simulator to analyze short-term problems, but at this time there is no simple way to transfer a DTA network to a micro-simulator. The process of manually developing an identical network in a traffic simulator can be time consuming and complex. For example, in a network developed in a microscopic simulator, vehicle rerouting is not captured; therefore, traffic flow in each link must be modified to reflect vehicle rerouting that occurs in DTA due to modifications, such as adding or dropping a freeway lane. As the network gets larger, more modifications to the traffic flow will occur as a result of the changing demands. Overall, researchers are limited to focus on either short-term or long-term situations.

Therefore, the goal for this study was to develop a program that builds an identical subnetwork in which vehicle routing data from VISTA is converted into traffic flows for CORSIM. Route data is converted to link flow data by gathering information about the number of vehicles whose paths cross the links per unit time.

7.1 Developing an identical network in CORSIM

This study will first show and explain the CORSIM input format, the tno and the trf files. Then, the next sections will discuss CORSIM required data, network information,

and traffic information. The first three sections are the network information, including the link details, such as link capacity, link length, the number of lanes, and free flow speed, the node positions, and the signal timing. The last two sections are the traffic information, including the entry flow and the turn movement percentages at each intersection.

CORSIM input format

There are two types of CORSIM input file, tno and trf files. The tno file is for CORSIM users to build the network manually by using the CORSIM simulators. Figure 51shows an example of tno file. When practitioners build a simulated network in a TNO file, it must be transformed to a digital file for CORSIM to read. On the other hand, a trf file can be transformed to a tno file if the practitioners want to review the simulated network. A digital trf file is shown below in Figure 52.

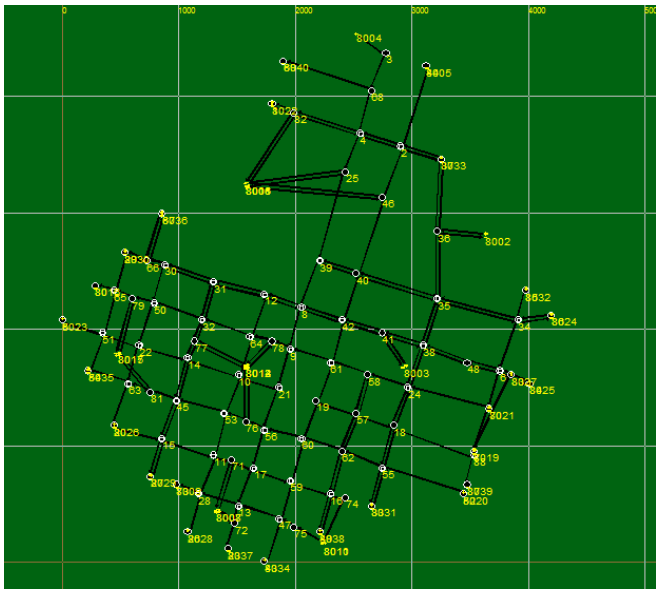


Figure 51 Example of CORSIM tno file

In a trf file, each row has 80 characters with meanings differentiated by Record Type (RT). The CORSIM Reference Manual describes all of the record types in detail. The record type includes the last three characters in each row. Error! Reference source not ound. is an example of the input with an explanation of the RT 11 entry codes. RT 11 defines the link description, including the up and down node, grade, distance, etc.

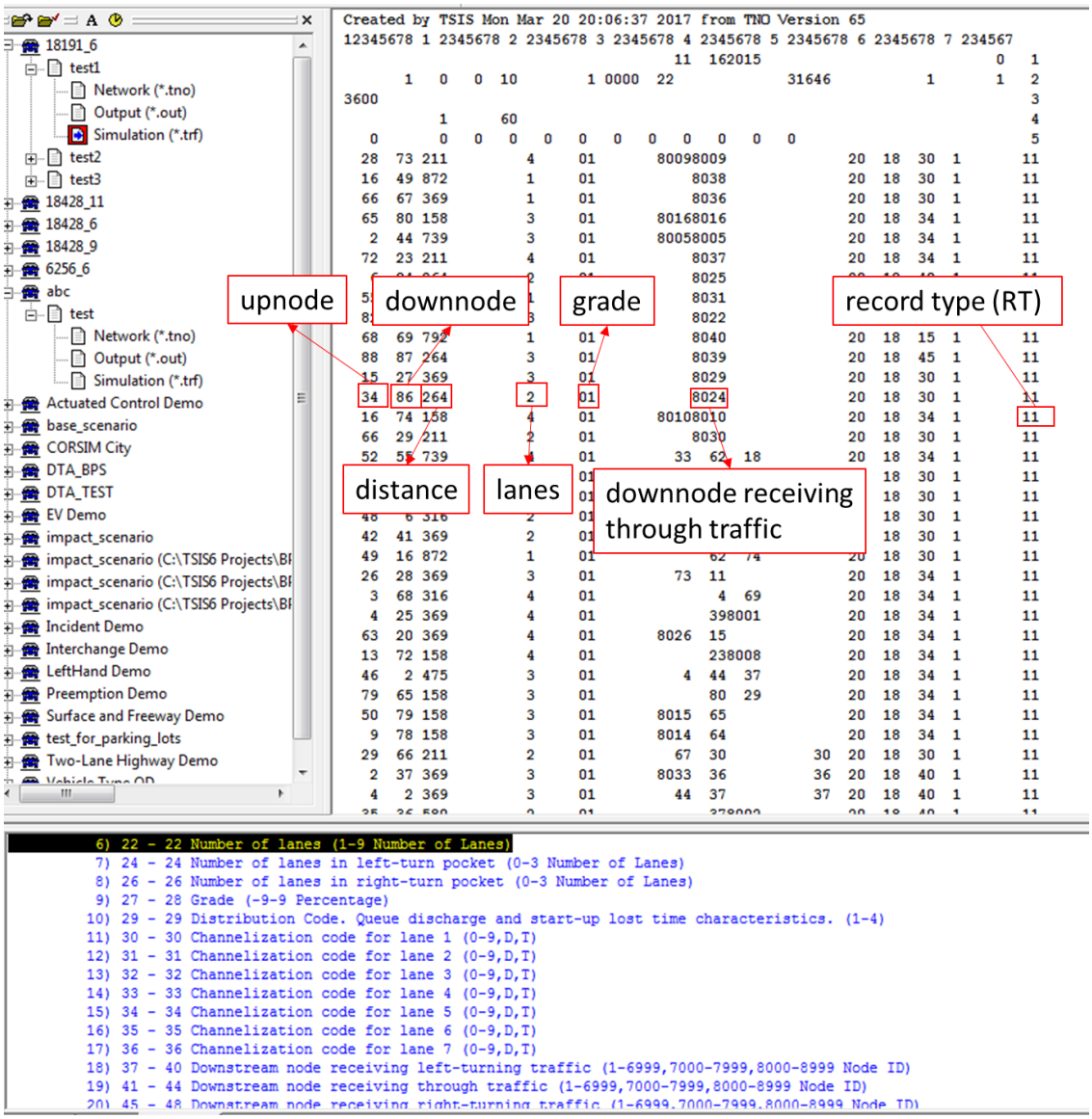


Figure 52 Example of CORSIM trf file

Network Information: Node Position

CORSIM's RT 195 describes the node position, including node id, x and y coordinates, whose units are in feet. VISTA provides the node position data (Table 18), and the columns in the table correspond to node id, longitude and latitude. The position units are not identical in CORSIM and VISTA, so the program used Equation (24) and (25) for transforming the position unit.

Table 18 Example of node position data from VISTA

ID	x	y
5106	-97.7388	30.2788
5107	-97.7395	30.2769
5135	-97.7514	30.2769

$$crsm\_node\_x = int((vsta\_node\_x - x) * 315634.7194176) + 1 \quad (24)$$

$$crsm\_node\_y = int((vsta\_node\_y - y) * 363692.5642944) + 1 \quad (25)$$

Where, crsm_node_x and crsm_node_y are the CORSIM node positions, and vsta_node_x and vsta_node_y are the VISTA node positions.

Network Information: Link Description

CORSIM's RT 11 (Figure 52) describes the link information, including upstream and downstream nodes, the link length, the number of lanes, free flow speed, and the downstream node receiving the direction of traffic, such as left turn, right turn, or through. Table 19 shows the link information from VISTA. The type includes 1 and 100, which are roads and connectors. CORSIM differentiates the links by the upstream and downstream node, which are called source and destination in Table 19. Also, the link type in CORSIM is differentiated by the upstream node. If a link upstream node is between 8000 and 8999, the link is defined as the boundary link in CORSIM.

Table 19 Example of link detail data from VISTA

ID	Type	Source	Destination	Length (ft)	Speed (mile/min)	Capacity (veh/hr)	Lanes
204866	100	100403	5571	475.2	1	800	1
304866	100	5571	200403	475.2	1	800	1
118380	1	13110	13109	369.6	0.5	800	1
18282	1	13026	13157	211.2	0.5	800	1

Most the link information can be obtained in VISTA link details, with the exception of

the downstream departure down nodes. VISTA does not contain turn movement information, which CORSIM needs in order to identify which downstream node should be a left turn node, a through node, or a right turn. Figure 53 shows the upstream, downstream, and downstream departure nodes in CORSIM.

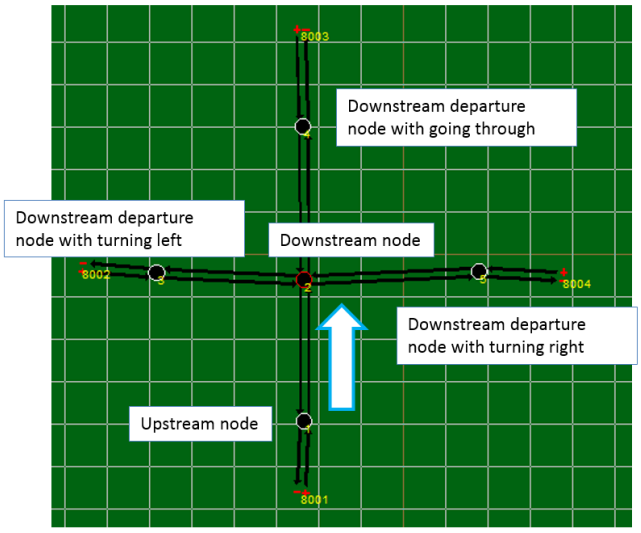


Figure 53 Example of upstream, downstream, and downstream departure nodes in CORSIM

To identify the direction of downstream departure down nodes, the system needs to define the orientation with respect to the node position. The process of identifying the node orientation is as follows:

1. Develop an N-by-N CORSIM link matrix; N is the number of connected nodes. The row and the column are the node id. The labeling of each cell is based on whether there is a link from the row's node to the column's node. If there is a link from node a to node b, this cell shows 1; otherwise, this cell shows 0. Table 20 is an example of the CORSIM link matrix. Because there are no links upstream and the downstream nodes are identical, the cells are 0 when the rows and the column nodes are identical. In the table, there is a two-way link between node 1 and node 3, but there is a one-direction link from node 3 to node 5.
2. Choose a cell whose value is one in the matrix: for example, (3, 5), the yellow background in Table 20.
3. Treat the cell's downstream node as the upstream node, and search for cells with a value of 1 and upstream node of 5. These cells, (5, 1) and (5, 6), the green

background in Table 20, are the downstream departure links.

4. The turning movements (left, through, right, or right diagonal) can be determined from ranking the clockwise angles measured from the upstream link.
    - If there are four downstream links, the link with minimum angle turns left, the next link goes through, the next links turns right, and the link with maximum is right diagonal.
    - If there are three downstream links, the links with minimum, medium, and maximum angle are left, through, and right.
    - If there are two downstream links, the link with the smallest clockwise angle turns left and the next link turns right.
    - If there is only one downstream link, the link serves only through (straight) traffic.
5. Restart at Step 2. The loop stops when all the cells in the matrix are identified.

Table 20 Example of the N-by-N connected node matrix

Upstream node id		Downstream node id						
		1	2	3	4	5	6	7
	1	0	0	1	0	0	0	0
	2	0	0	0	0	0	0	1
	3	1	0	0	0	1	1	0
	4	0	0	0	0	0	0	0
	5	1	0	0	0	0	1	0
	6	0	1	0	0	0	0	0
	7	0	0	1	0	1	0	0

Traffic Information: Entry Flow

CORSIM's RT 50 defines the entry link and flow, and RT 51 defines the traffic volumes on source/sink links. The difference between RT 50 and 51 is that the former is for flows at the boundaries of the network, while the latter is for flows within the network. RT 50 defines the entry flow from the boundary of the network (blue circle in Figure 54 and Figure 55), and RT 51 is the entry flow in the network, similar to the parking lots (green circle in Figure 54 and Figure 55). Figure 55 is the CORSIM network in animated form. The boundary links and nodes are virtual and are not represented in the

animation, but the source/sink links are shown in the animation. In contrast, VISTA has two kinds of entry flow: one from the boundary nodes (the green nodes with the black links in Figure 56), and the other from centroids (the green nodes with gray links in Figure 56). To represent these different types of entry flow in CORSIM, RT 50 is assumed to represent the boundary flow specifications and RT 51 as the flows entering the network from traffic sources inside the network (for example parking garages).

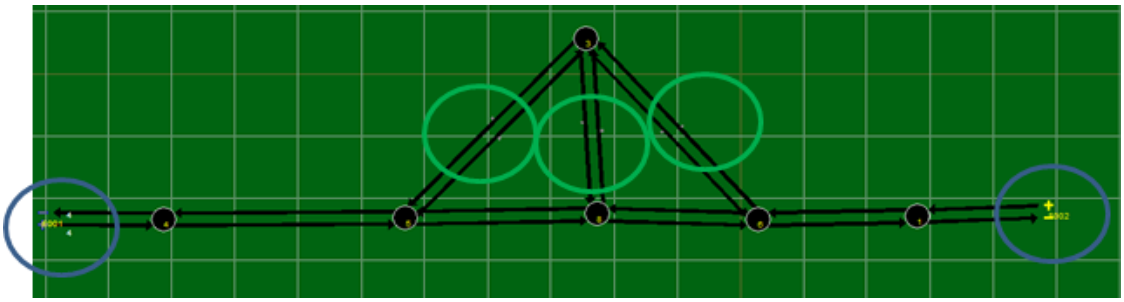


Figure 54 Example of CORSIM's entry flow from the boundary or inside of the network

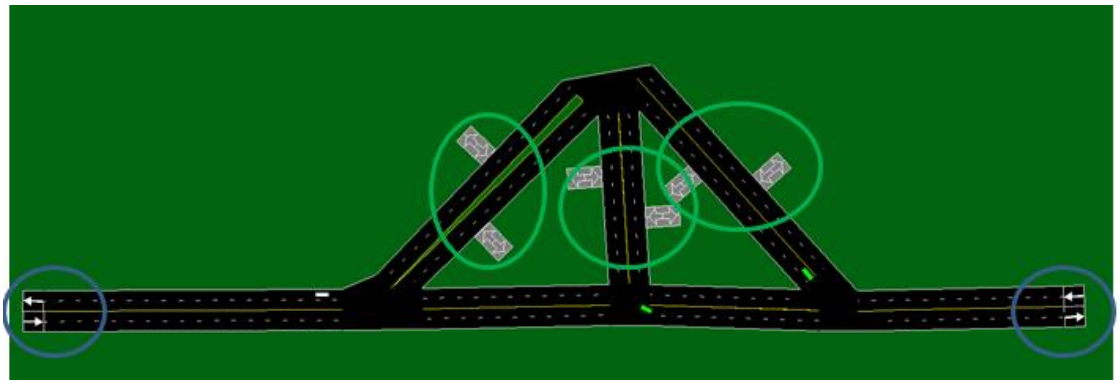


Figure 55 Example of the CORSIM network in animated form

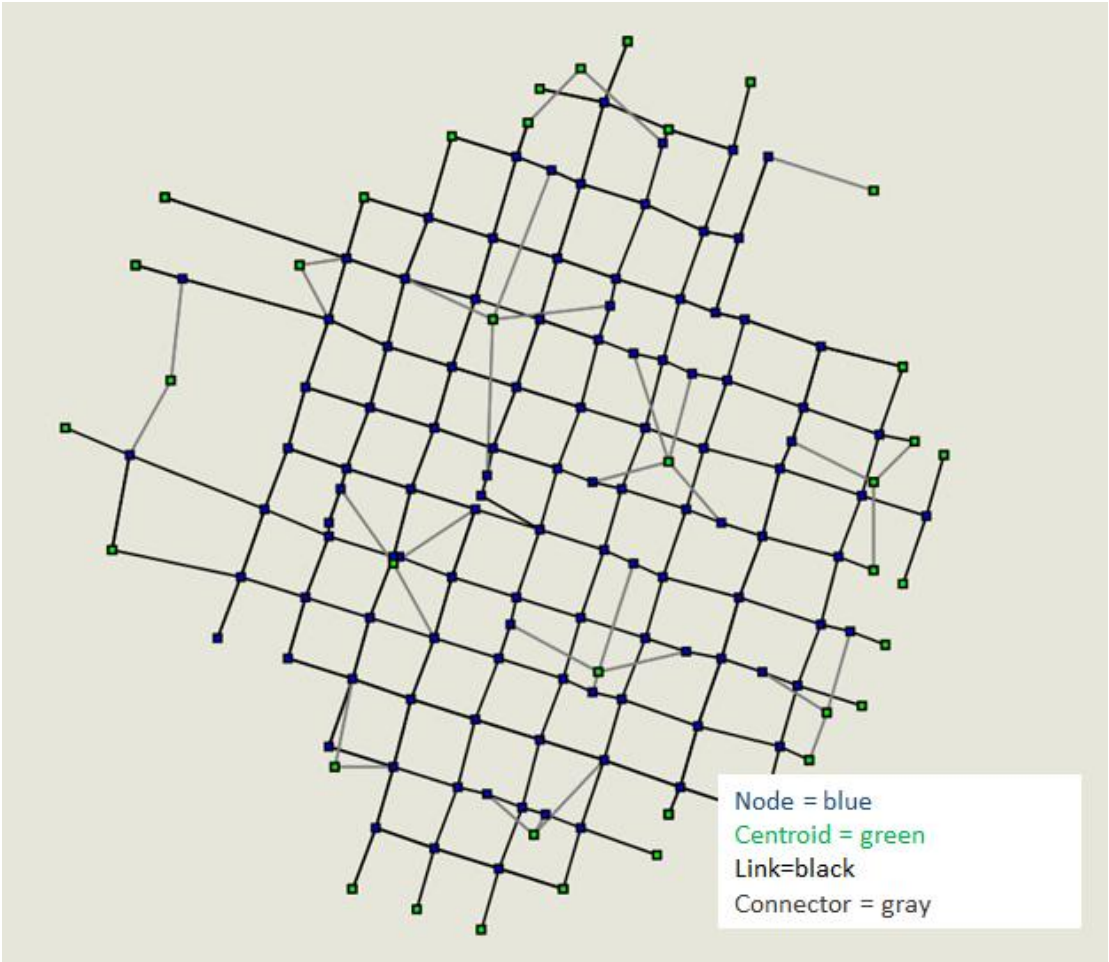


Figure 56 Example of centroids and connectors in VISTA

However, RT 51 has a restriction: the number describing the exiting flow cannot be more than 999 veh /hr. It is easy to have more traffic flow in a congested area, such as the downtown areas. If the traffic flow is more than 999 veh/hr, the simulated network cannot be implemented. Conversely, RT 50 does not have this limitation on exiting traffic flow, so for the purpose of this study, the researchers decided to only use RT 50 for the boundary and connector flows.

Figure 57 shows the example of the required entry code in RT 50, the boundary link's upnode and down node, the entry flow, and RT number. The boundary link's upnode number should be between 8000 and 8999.

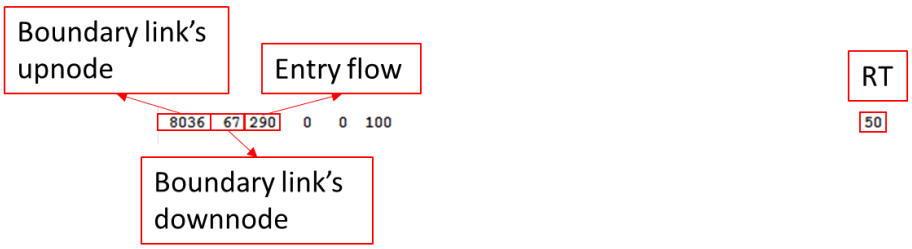


Figure 57 Example of RT 50 in trf files

VISTA provides vehicle-tracking data after each simulation, including vehicle path and path time. Vehicle path data provides path IDs, and each path ID contains the link order from the origin link ID to the destination link ID. Also, vehicle path time data includes each vehicle path ID and the entering times to link ID's in the path. Table 21 displays the vehicle path data; *id* means the vehicle ID number, *Origin* and *Destination* are the departure and destination link ID numbers, *freeflowtt* is the free flow travel time between the origin and destination, *length* is the sum of the link lengths in feet, and *Link ID Numbers* represents the link ids included in the path. Table 22 displays the vehicle path time data where *sim_departure* is the vehicle departure time, *sim_exittime* is the time the vehicle arrives at the destination, *dta_path* is the path ID (also in Table 22) that the vehicle chooses corresponding to the vehicle path data, and *arrival time to each link* is the arrival time for the vehicle to get to each link in that path. For example, the first line in Table 21 corresponds to Path 1 and the Link ID numbers in the path are listed in the last column. In Table 22, the first line corresponds to a specific vehicle on Path 1 and the last column shows the arrival time, in seconds, to the links listed under Path 1 in Table 21. Vehicle 10285 in Table 22 chooses path 1 (link ID orders: 218517, 116214, 18363, 18557, 18364, 18365, 363101) in Table 21. This vehicle arrives at the first link (218517) in 4013 seconds, the second link (116214) in 4014 seconds, etc.

Table 21 Example of vehicle path data

Path ID	*Origin (link id)*	*Destination (link id)*	*freeflowtt (sec)*	*Length (ft)*	*Link ID Numbers*
1	100362	263101	222	7480	{218517,116214,18363,18557,18364,18365,363101}
2	100356	200360	12	616	{206242,6237,318525}
3	156336	200361	126	7392	{256336,115778,106199,118516,116214,106244,115777,118528,106236,118532,318533}

Table 22 Example of vehicle path time data

Vehicle ID	*sim_departure (sec)*	*sim_exittime (sec)*	*dta_path (path id)*	*Arrival time to each link (sec)*
10285	4013	4125	1	{4013,4014,4020,4041,4059,4098,4125}
10286	4210	4329	1	{4210,4215,4221,4242,4260,4302,4329}
10287	4355	4467	1	{4355,4356,4362,4383,4401,4440,4467}

Using the vehicle path data and the vehicle path time data, counting the number of vehicles that pass across the link in a given time period calculates the traffic flow on each link. This information provides the boundary flow and the turn movement. Although the CORSIM boundary links cannot share the same boundary node, a VISTA centroid usually has several connectors (Figure 58). Thus, the program generates multi-boundary nodes in CORSIM to replace a centroid in VISTA. The multi-boundary nodes are at the identical position but have different id numbers. Each boundary node connects

to the connector's node. Figure 59 shows an example of the multi-boundary nodes in CORSIM from a VISTA centroid. The left side is from CORSIM and the right is from VISTA. The red circles show the boundary links on the left side and connectors on the right side. In the blue circles, there are multi-boundary nodes on the left side and only one centroid on the right side. Each boundary node connects a connected node in left side, and only one centroid connects to distinct connected nodes in the right side. Then, the auto program can define the boundary flows by using the connector flows.

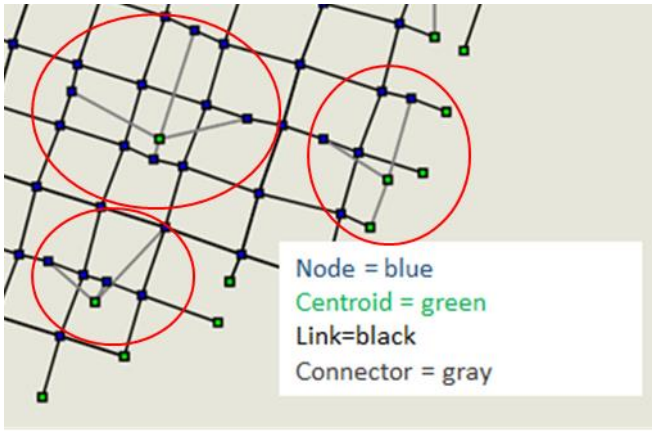


Figure 58 Example of multi-connectors to a centroid in VISTA

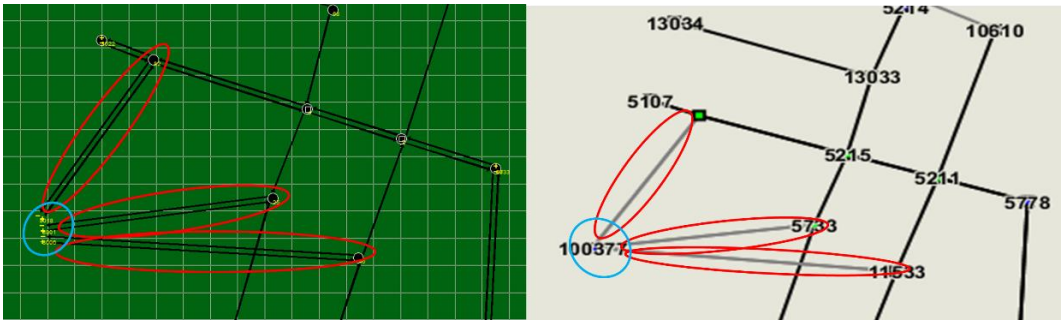


Figure 59 Example of multi-boundary nodes in CORSIM from a VISTA centroid

Traffic Information: Turn Movement

CORSIM's RT 21 defines the upstream link's turn movement. The second row in Figure 60 shows the entry codes in RT 21 in CORSIM. The first two numbers correspond to the link upnode and down node, and the next three numbers are the turn movement from turn left, through, to turn right. The unit of the turn movement corresponds to traffic flow (veh/hr) or ratio. Although VISTA does not provide the turn movement, the auto program defines the departure down nodes by using the upstream

and downstream angles in RT 11 (first row in Figure 60). Also, each link flow from upstream to downstream is defined by integrating the vehicle path and vehicle path time. Thus, the auto program can provide the traffic turns by using the downstream departure nodes. For example, in Figure 60, RT 11 shows that the upstream link's downstream departure nodes are node 41, 35, and 48 for turning left, through, and right. The program knows that the traffic from upstream links to node 41, bn35, and 48 are is 76 veh/hr, 174 veh/hr, and 17 veh/hr, so the left turn, through, and right turn traffic in RT 21 are 76 veh/hr, 174 veh/hr, and 17 veh/hr.

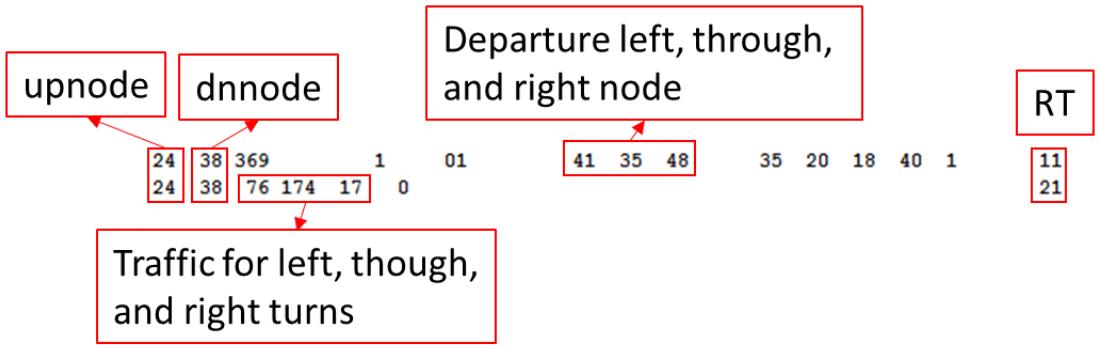


Figure 60 Example of RT 11 and RT 21 in trf files

Network Information: signal timing

CORSIM's RT 35 and RT 36 define the signal timing. RT 35 provides the intersection node's id, the upstream node to this intersection node, and the signal timing. RT 36 provides the signal code for each phase, and the order of the signal code is from the order of the upstream approach nodes. Also, Figure 61 shows CORSIM sign and pre-timed signal control codes. In Figure 62, the first two signal codes are 2 and 1, which means that node 51 is a red light and node 79 is a green light for 29 seconds. The next two signal codes are 2 and 0, which means that node 51 is still a red light, but node 79 turns to a yellow light for five seconds.

Table 23 provides the time offset data in VISTA. The node ID is the intersection node, and the offset is the offset time.

Table 24 shows the signal timing data from VISTA. Signal ID is the intersection node id * 100 + the phase order. Phase shows the phase order. Red, Yellow, and Green show all of the red, the yellow, and green phase times. Each value in the upstream and downstream links are the approving path from the upstream to the downstream. For

example, at Signal ID 1120601, the approving path is from link id 18492 to link id 118216 and 5244, and from link id 105244 to 118216 and 118492.

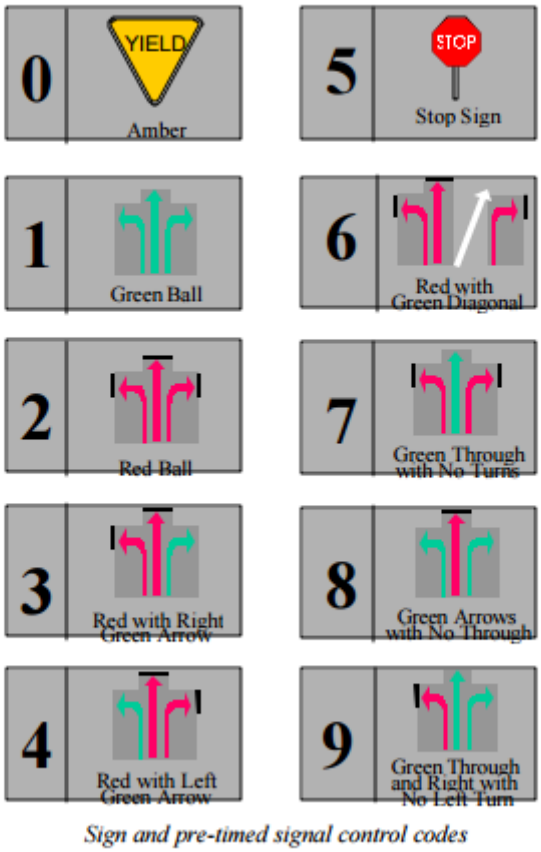


Figure 61 CORSIM's sign and pre-timed signal control codes

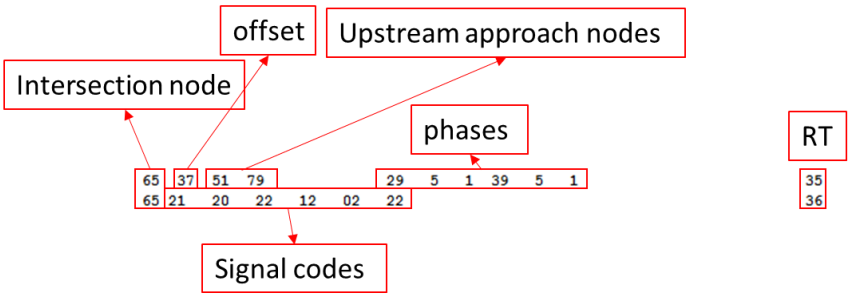


Figure 62 Example of the RT 35 and RT 36 entry codes in the trf file

Table 23 Example of signal offset data from VISTA

Node ID	Time offset (sec)
5458	0
13075	0
5758	66
5759	17

Table 24 Example of signal timing data from VISTA

Signal ID	Node ID	Phase	Red (sec)	Yellow (sec)	Green (sec)	Upstream link	Downstream link
1120601	11206	1	1	3	62	{18492,18492, 105244,105244}	{118216,5244, 118216,118492}
1120602	11206	2	1	3	50	{18216,18216}	{118492,5244}

Some signals are more complicated in the real world. For one approach (or one link) to the intersection, it could be green, permitting a left turn in the first phase, then become green for all possible travel directions (left, right, and through) in the second phase.

Table 24 shows that the signals in VISTA represent the green light by showing which upstream and downstream links are green rather than showing the signal light. Thus, VISTA does not contain information about traffic direction, such as turning left, through, or right.

The traffic signal is represented in CORSIM in a manner similar to the real world. The system determines the downstream link direction/orientation based on the *Turn*

Movement, and uses a signal light to represent which downstream links are allowed to proceed. The CORSIM transformer performs the following steps to convert the signal lights from VISTA to CORSIM:

1. Collect the approved downstream links for each upstream link.
2. For each upstream link, search for the downstream links in *Turn movement* that are not included in the downstream links in step 1. The included downstream links are green lights and the absent downstream links are red lights.
3. Choose the right signal code by searching Figure 61.

7.2 Result

This study chose data taken at the intersection of Guadalupe Street and 24th Street to extract a subnetwork in VISTA (Left figure in Figure 63), and used the CORSIM transformer to develop the same network (Right figure in Figure 63). From the geometry side, both networks are similar.

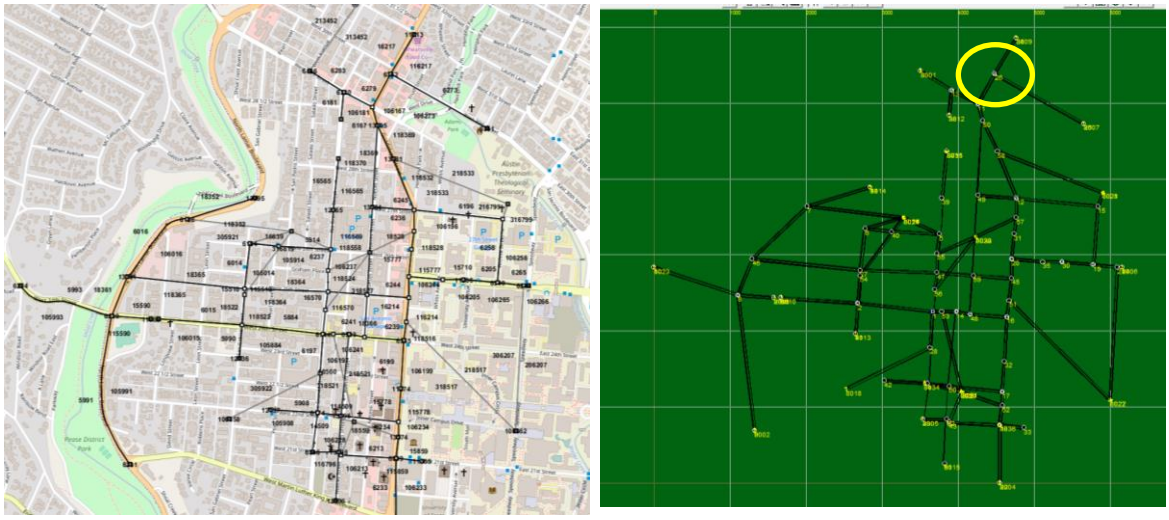


Figure 63 The transformed network by using CORSIM Transformer

However, Table 25 shows the path travel time where the difference is greater than 50% in VISTA and CORSIM. In Table 25, Path is the vehicle path id in VISTA. VISTA and CORSIM show the VISTA and CORSIM's path travel time, and % shows the ratio of the difference to the VISTA's path travel time. The bold number is the ratio greater than 100% or less than -100%.

The main reason for the differences is that the vehicles in VISTA do not have driving behavior whereas CORSIM vehicles do have driving behavior. The CORSIM left-

turning or right-turning vehicles are programmed to hesitate and wait. Since these vehicle are stopped, the other vehicle behind them are also stopped, creating a long line of paused vehicles. Therefore, the network from the CORSIM Transformer still requires more calibration, but the advantage is that CORSIM converts the paths to the traffic flow easily. When the network is calibrated and the paths are changed from VISTA, the user can simply change the traffic flow by using the CORSIM Transformer rather than checking each turn movement and entry flow at all nodes manually.

Table 25 The ratio of travel time difference between VISTA and CORSIM

Path	Vista	CORSIM	%
4	100	162	-62
7	87	37	57
10	**94**	**243**	**-159**
11	**70**	**260**	**-271**
16	21	38	-81
17	**17**	**299**	**-1659**
22	**115**	**442**	**-284**
46	282	122	57
48	20	38	-90
54	**98**	**243**	**-148**
59	**109**	**410**	**-276**
80	9	4	56
96	**263**	**617**	**-135**
101	**161**	**370**	**-130**
105	31	51	-65
110	**79**	**341**	**-332**
112	**139**	**407**	**-193**
115	**186**	**450**	**-142**
123	33	51	-55
129	**160**	**426**	**-166**
130	45	69	-53
138	19	35	-84
150	105	174	-66
180	129	11	91

7.3 Summary

This study developed a program that automatically converts an identical CORSIM subnetwork from VISTA subnetwork data. The required CORSIM input includes link description, node position, signal timing, entry flow, and turn movement data. In addition, there are CORSIM input formats: tno and trf files. The tno file is built

manually in CORSIM, and the trf file is a digital file. This program developed the trf file by using the VISTA link details table, the node position table, the signal timing and offset table, and the vehicle path and vehicle path time table.

While the concepts of building a network in CORSIM and in VISTA are similar, there are three key differences. First, the network in VISTA does not have turn movement information, which CORSIM requires in order to identify which downstream node should be a left turn node, a through node, or a right turn node. This study used the angle of the upstream and downstream link to define downstream orientation; the smallest is a left turn, the middle is a through, and the largest is a right turn for a three-leg intersection. Then, the number of vehicles was determined by integrating the vehicle path and vehicle path time data. Finally, the orientation of downstream departure nodes and the number of vehicles to the departure nodes revealed the turn movements.

Second, CORSIM boundary nodes cannot connect to more than one connected node, but a VISTA centroid usually has more than connector. The automatic program generates multi-boundary nodes at the same position from the position of the VISTA centroid. The CORSIM boundary flows from a boundary node and a connected link is generated from the connector flow in VISTA.

Third, the completed CORSIM subnetwork should be calibrated manually by using CORSIM transformer due to vehicle behavior. However, even though the network should be checked, the user should check their network if it was built manually. However, the user does not need to check every turn movement and entry flow at all nodes manually if the network is calibrated, since CORSIM transformer can change the flow easily.

CHAPTER 8 FINAL

8.1 Summary

This dissertation aimed to develop robust tools that enable efficient and comprehensive network analysis for both a VISTA (DTA) and CORSIM (microsimulation) setting. The tools built include 1) a definition for an optimal sized subnetwork that was implemented into a program that automatically extracts an optimal sized subnetwork from a regional network and 2) another program that automatically converts the DTA subnetwork into a CORSIM format.

Using a sub-area to represent the entire network performance is one way to reduce computational burden. Bringardener (2014) developed two models that defined the minimum sub-area size for the user defined acceptable error. However, there is no clear-cut way for a user to define the error. Additionally, simulation time and acceptable error do not change proportionally. This study found several parameters can affect simulation time, including the number of vehicles (demand), the number of links (link), the sub-area order (order), and the work zone sizes (zone) to better characterize how simulation time can change.

Most of the tested parameters are correlated to each other, except for zone and order. This study found that demand is most correlated to simulation time (Adjusted R square = 0.65). Also, this study used zone and size to build a linear regression model (Adjusted R square = 0.603) to predict simulation time. Having the demand data requires extracting a sub-area first. Demand is most correlated to simulation time, but if the user would like to predict the approximate simulation time without having created a sub-area, then using the second model to estimate the simulation time also works well. It was outside the scope of this study to generalize simulation time in terms of clock speeds or other potential general measures.

Additionally, this study developed two models, a confidence interval model and an impact model, that can define the sub-area size that has 95% significant impact to identify whether there is an impact on a link. The confidence interval model estimates the target link's RMSE range in the base scenario (called base RMSE), and the impact model estimates the target link's RMSE value in the impacted scenario (called impact RMSE). If the base RMSE is larger than the impact RMSE, then the target link is not

impacted by the work zone. Conversely, if the base RMSE is less than the impact RMSE, the target link has a 95% significant impact from the work zone.

Also, this study developed two objective functions to define the balance between the simulation time and error associated with a sub-area. The functions defined the sub-area size and the computational time by using Bringardner (2014) RMSE models and the estimated simulation time model using zone size , sub-area size, and zone average flow, and the independent variables. Both functions provide users a way of defining the optimal sub-area size.

Furthermore, while the results from a DTA network are typically similar to what is expected in the real-world, DTA results do not have as many details as those coming from a microsimulator. By converting a DTA network into a CORSIM-compatible form, the analyst can easily extract the necessary detailed information to conduct operational assessments. This study developed two programs for extracting a DTA sub-area from VISTA (called the DTA Sub-area Developer) and transforming it into a CORSIM compatible format (called the CORSIM Transformer). Users do not need to learn the process of extracting a sub-area if using the DTA sub-area developer. Also, users can have the identical CORSIM network from a VISTA network when using both programs.

Ultimately, these programs will attract researchers and practitioners to use DTA more than before, study sub-areas, and analyze traffic conditions more efficiently and completely. These programs will enable users to account for the path changing that occurs as a result of a work zone or other change in the network without having to worry about the computational burden nor sacrificing accuracy of the results. The output of these simulations can be easily converted into CORSIM format with the CORSIM Transformer program to facilitate traffic operations analysis.

8.2 Contribution

This contribution from this study includes intellectual and practical contribution. The intellectual contribution is as following,

1. This study developed indexes (demand, link, zone size, sub-area size) to estimate the computational time. Users can define the optimal sub-area size without having to choose tolerable error.
2. This study developed VISTA sub-area extractor, which removes the need for manual steps prone to error for treating link changes as network problems rather than local problems.
3. This study developed CORSIM sub-area transformer, which facilitates the transition from a mesoscopic to a microscopic simulation. Also, users can have highly detailed analysis, and describes network more effectively by using CORSIM simulator.

The practical contribution is as following,

1. The automatic programs and tools developed open the door for more people to benefit from DTA sub-network analysis by making the tool much easier to use and overall more accessible.
2. Since the computational time will be reduced by using the automatic tools and models, practitioners can evaluate multiple scenarios more easily and the optimal alternatives to solve a variety of network problems, i.e. work zones, or building a new freeway
3. In the past, VISTA was a tool for long term analysis and CORSIM was for short term. This study used VISTA to assign the traffic and CORSIM to analyze the traffic conditions. Practitioners and officials can use these tools to decide on optimal alternatives according to both short-term and long-term analysis.

8.3 Future Research

The future research can estimate the sub-area size when the work zones are not connected. This study estimated the sub-area sizes when the work zones are connected. The subnetwork size may be different when the work zones are at different locations. Some cities are in the development, and these cities may have different work zones simultaneously. The estimated sub-area size from this research may not applicable in these cities.

The future research can consider the traffic reassignment when the OD matrixes has changed. This study did not consider multi-transportation tools, and the users may

change their transportation modes if the traffic condition is severe. The OD matrixes change if some drivers shift to using the public transport.

APPENDIX

```
//Code for extracting sub-area's links and nodes
for (int i = 1; i <= order + 1  ; i ++)
{
    if (i ==2 )
        int abc=123;
    for (sublk_pos= sublk.begin(); sublk_pos != sublk.end();
sublk_pos++ )
    {
        int id1 =  (*sublk_pos).ID();
        int org1 = (*sublk_pos).DEST();
        int org2 = (*sublk_pos).ORG();
        flag = (*sublk_pos).FLAG();
        float distance1 =  (*sublk_pos).returnDistance();
        float length1 = (*sublk_pos).returnLength();
        float total_distance;
        if (i == 1) total_distance = 0;
        else total_distance = distance1+length1;

        if (i < order +1 )
        {
            for (lk_pos = lklist.begin();  lk_pos !=lklist.end();
lk_pos++)
            {
                int id = (*lk_pos).ID();
                int org = (*lk_pos).ORG();
                int dest = (*lk_pos).DEST();
                if ((*lk_pos).FLAG()== 1) continue;
                if ( (*lk_pos).TYPE() == 100) continue;

                if ((*lk_pos).ORG()== org1 || (*lk_pos).ORG() ==
org2 || (*lk_pos).DEST() == org1 || (*lk_pos).DEST()==org2)
                {
                    (*lk_pos).details((*lk_pos).ID(),
(*lk_pos).ORG(), (*lk_pos).DEST(), (*lk_pos).TYPE(),
1,(*lk_pos).returnSpeed(),
(*lk_pos).returnCapacity(),(*lk_pos).returnLanes() );

                    int sub_flag = i; //parameter order
                    linklist lk;
                    lk.details((*lk_pos).ID(), (*lk_pos).ORG(),
(*lk_pos).DEST(),(*lk_pos).TYPE(), sub_flag,
(*lk_pos).returnSpeed(),
(*lk_pos).returnCapacity(),(*lk_pos).returnLanes() );
                    lk.inputLength((*lk_pos).returnLength());
                    lk.InputDistance(total_distance);

                    temlk.push_back(lk);
                    continue;
                }
            }
        }
        else if (i == order + 1)
        {
            for (lk_pos = lklist.begin();  lk_pos !=lklist.end();
lk_pos++)
            {
                if ((*lk_pos).FLAG()== 1) continue;
                if ( (*lk_pos).TYPE() == 1) continue;
```

```
                if ((*lk_pos).DEST()== org1 || (*lk_pos).ORG() ==
org2 || (*lk_pos).DEST() == org2 || (*lk_pos).ORG()==org1)
                {
                    (*lk_pos).details((*lk_pos).ID(),
(*lk_pos).ORG(), (*lk_pos).DEST(), (*lk_pos).TYPE(),1,
(*lk_pos).returnSpeed(),
(*lk_pos).returnCapacity(),(*lk_pos).returnLanes() ); //flag = 1
read

                    int sub_flag = i; //parameter order

                    linklist lk;
                    lk.details((*lk_pos).ID(), (*lk_pos).ORG(),
(*lk_pos).DEST(), (*lk_pos).TYPE(),sub_flag,
(*lk_pos).returnSpeed(),
(*lk_pos).returnCapacity(),(*lk_pos).returnLanes() );
                    lk.inputLength((*lk_pos).returnLength());
                    lk.InputDistance(total_distance);
                    temlk.push_front(lk);
                    continue;
                }
            }
        }
    }
    for( lk_pos = temlk.begin(); lk_pos != temlk.end(); lk_pos++)
    {
        int sub_id = (*lk_pos).ID();
        int sub_org = (*lk_pos).ORG();
        int sub_dest = (*lk_pos).DEST();
        //int sub_flag = order; //parameter order

        int sub_flag = (*lk_pos).FLAG(); //parameter order
        float distance = (*lk_pos).returnDistance();
        float length = (*lk_pos).returnLength();
        linklist lk;
        lk.details((*lk_pos).ID(), (*lk_pos).ORG(),
(*lk_pos).DEST(),(*lk_pos).TYPE(), sub_flag,
(*lk_pos).returnSpeed(),
(*lk_pos).returnCapacity(),(*lk_pos).returnLanes() );
        lk.inputLength(length);
        lk.InputDistance(distance);
        sublk.push_back(lk);
    }
    temlk.clear();
}
```

Figure 64 Code for extracting sub-area's links and nodes from an entire network (Language: C++).

```
for (int i = 0; i < network_size; i ++)
{
    vista_network[i].resize(network_size);
    for (int j = 0; j < network_size; j ++)
    {
        vista_network[i][j]=0;
        if (i==j) continue;
        if (vista_links[i][4]==vista_links[j][3] &&
vista_links[i][3]!=vista_links[j][4])   vista_network[i][j]=1;
        if (vista_links[i][4]==vista_links[j][3] &&
vista_links[i][3] ==vista_links[j][4])
        {
            int times = 0;
            for( lk_pos = linklist.begin(); lk_pos !=
linklist.end(); lk_pos++) //find upstream's orientation
            {
                int linkid = (*lk_pos).returnVistaID();
                if (linkid == vista_links[i][0])
                {
                    times++;
                    (*lk_pos).Twoways( vista_links[j][0]); //to
record the other side of the link
                }
                else if (linkid == vista_links[j][0])
                {
                    times++;
                    (*lk_pos).Twoways( vista_links[i][0]); //to
record the other side of the link
                }
                if (times== 2) break;
            }
        }
    }
}
```

Figure 65 Code for defining the subnetwork in CORSIM (Language: C++).

```
for (path_pos = vsta_pathlist.begin(); path_pos!=
vsta_pathlist.end(); path_pos++)
{
    int pathid = (*path_pos).returnPathID();
    vector <int> links = (*path_pos).returnLinklist();
    vector <int> vehs = (*path_pos).returnVehIDs();
    vector <vector <int>>vehtimes = (*path_pos).returnVehTimes();
    cout <<pathid <<endl;
    for (int i = 0; i < links.size() - 1; i ++)
    {
        int upstream = links[i];
        int dnstream = links[i+1];
        if (upstream ==6284)
            int abc = 123;
        if (dnstream == 6167)
            int abc = 123;
        for (lk_pos = linklist.begin(); lk_pos!= linklist.end();
lk_pos++)
        {
            int linkid = (*lk_pos).returnVistaID();
```

```
            if (linkid ==upstream || (i+1 == links.size() - 1 &&
linkid ==dnstream))
            {
                for (int j = 0; j < vehs.size(); j ++)
                {
                    vector <int> inf;
                    inf.push_back(dnstream);
                    inf.push_back(vehs[j]);
                    inf.push_back(vehtimes[j][i+1]);
                    (*lk_pos).DnsVehs(inf);
                }
            }
        }
    }
}
```

Figure 66 Code for transforming from vehicle paths to link volumes (Language: C++).

Parameters

C Link Capacity /6000/
TC Target link capacity /6000/
v_over_c V over C /0.159/
R Capacity Reduction /25/
N Zone Size /1/;

Positive VARIABLES

Perfect_Size network size without error
Optimal_Size optimal network size considering simulation time and error ;

VARIABLES

Z ;

EQUATIONS

NO_ERROR network size without error
Define_Perfect_Impact_Regression
Limitation
Optimal ;

Optimal..Z=E=1-Optimal_Size/Perfect_Size+(-20.075+31.823*Optimal_Size+17.984*N)/(-20.075+31.823*Perfect_Size+17.984*N) ;

NO_ERROR .. -0.115*Perfect_Size -0.018*C/1000-15.487*v_over_c+6.004 =E= 0.430*Perfect_Size +0.135*R + 0.719*N-0.117/1000 * C -9.610*v_over_c-0.016*Perfect_Size*R ;

Define_Perfect_Impact_Regression .. Optimal_Size =L=Perfect_Size;

Limitation .. Optimal_Size =G=0;

```
MODEL m /ALL/ ;
SOLVE m USING NLP Minimize Z;
```

Figure 67 Code for optimizing sub-area size (Language: GAMS)

REFERENCES

Bloomberg, L., and Dale, J. (2000). A comparison of the VISSIM and CORSIM traffic simulation models. In Institute of Transportation Engineers Annual Meeting.

Bringardner, J. W., Gemar, M. D., Boyles, S. D., and Machemehl, R. B. (2014). Establishing the Variation of Dynamic Traffic Assignment Results Using Subnetwork Origin-Destination Matrices. In Transportation Research Board 93rd Annual Meeting (No. 14-3107).

Bringardner, J. W. (2015). Application of a subnetwork characterization methodology for dynamic traffic assignment. (Unpublished doctoral dissertation). University of Texas at Austin.

Chen, B. Y., Lam, W. H., Sumalee, A., Li, Q., and Li, Z. C. (2012). Vulnerability analysis for large-scale and congested road networks with demand uncertainty. Transportation Research Part A: Policy and Practice, 46 (3), 501-516.

Chiu, Y. C., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Waller, T., & Hicks, J. (2011). Dynamic traffic assignment: A primer. Transportation Research E-Circular, (E-C153).

Choa, F., Milam, R. T., and Stanek, D. (2004). Corsim, paramics, and vissim: What the manuals never told you. In Ninth TRB Conference on the Application of Transportation Planning Methods.

Gemar, M. D. (2013). Subnetwork analysis for dynamic traffic assignment: Methodology and application (Unpublished doctoral dissertation). University of Texas at Austin.

Gemar, M., Bringardner, J., Boyles, S., and Machemehl, R. (2014). Subnetwork Analysis for Dynamic Traffic Assignment Models: Strategy for Estimating Demand at Subnetwork Boundaries. Transportation Research Record: Journal of the Transportation Research Board, (2466), 153-161.

Gomes, G., May, A., and Horowitz, R. (2004). Calibration of VISSIM for a Congested Freeway. California Partners for Advanced Transit and Highways (PATH).

Horowitz, R., May, A., Skabardonis, A., Varaiya, P., Zhang, M., Gomes, G., and Sun, D. (2005). Design, field implementation and evaluation of adaptive ramp metering algorithms. *California Partners for Advanced Transit and Highways (PATH)*.

ITT Industries, Inc., Systems Division ATMS R&D and Systems Engineering Program Team (2001). CORSIM Reference Manual. The attachment is available at http://www.chinautc.com/information/manage/UNCC_Editor/uploadfile/20081022155355673.pdf

ITT Industries, Inc., Systems Division ATMS R&D and Systems Engineering Program Team (2006a). CORSIM Data Dictionary. The attachment is available at http://sites.poli.usp.br/ptr/lemt/CORSIM/CORSIM%20Data%20Dictionary.pdf

ITT Industries, Inc., Systems Division ATMS R&D and Systems Engineering Program Team (2006b). CORSIM User's Guide. The attachment is available at http://sites.poli.usp.br/ptr/lemt/CORSIM/CORSIMUsersGuide.pdf

ITT Industries, Inc., Systems Division ATMS R&D and Systems Engineering Program Team (2006c). TSIS User's Guide. The attachment is available at http://www.et.byu.edu/~msaito/CE662MS/User%20manuals/TSIS/TSISUsersGuide.pdf

Larsson, T., Lundgren, J. T., Rydergren, C., & Patriksson, M. (2001). Most likely traffic equilibrium route flows analysis and computation. In Equilibrium Problems: Nonsmooth Optimization and Variational Inequality Models (pp. 129-159). Springer US.

Levin, M. W., Pool, M., Owens, T., Juri, N. R., & Waller, S. T. (2015). Improving the convergence of simulation-based dynamic traffic assignment methodologies. Networks and Spatial Economics, 15(3), 655-676.

Mathew, T. V., and Krishna Rao, K. V. (2006). Introduction to Transportation engineering. Civil Engineering–Transportation Engineering. IIT Bombay, NPTEL ONLINE. The attachment is available at http://www. cdeep. iitb. ac. in/nptel/Civil% 20Engineering.

Meyer, M. D., & Miller, E. J. (2001). Urban transportation planning: A decision-oriented approach. New York: McGraw-Hill.

McNally, M. G. (2007). The four step model. Handbook of transport modelling, 1, 35-41.

Röder, D., Cabrita, I., and Nagel, K. (2013). Simulation-based sketch planning, part III: Calibration of a MATSim-model for the greater Brussels area and investigation of a cordon pricing for the highway ring. VSP working paper 13-16, TU Berlin, Berlin, Germany.

Visual Solutions, Inc. (2010). VisSim User's Guide. The attachment is available at http://www.vissim.com/downloads/doc/VisSim_UGv80.pdf

Yperman, I. (2007). The link transmission model for dynamic network loading. The attachment is available at http://www.kuleuven.be/traffic/dwn/P2007A

Zhou, X., and Mahmassani, H. S. (2006). Dynamic origin-destination demand estimation using automatic vehicle identification data. IEEE Transactions on intelligent transportation systems, 7(1), 105-114.